International Journal of

Computational Linguistics & Chinese Language Processing

中文計算語言學期刊

A Publication of the Association for Computational Linguistics and Chinese Language Processing

This journal is included in THCI, Linguistics Abstracts, and ACL Anthology.

易繫辭曰上古結繩而治後世聖人易之以書契百官以治萬民以察說文敘曰蓋文字者經藝之本宣教明化之始前人所以垂後後人所以識古故曰本立而道生知天下之至賾而不可亂也教化既萌文心雕龍則謂人之立言因字而生句積句而成章積章而成篇篇之彪炳

Vol.23 No.1 June 2018 ISSN: 1027-376X

International Journal of Computational Linguistics & Chinese Language Processing

International Journal of

Computational Linguistics & Chinese Language Processing

Aims and Scope

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) is an international journal published by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP). This journal was founded in August 1996 and is published four issues per year since 2005. This journal covers all aspects related to computational linguistics and speech/text processing of all natural languages. Possible topics for manuscript submitted to the journal include, but are not limited to:

- Computational Linguistics
- Natural Language Processing
- Machine Translation
- Language Generation
- Language Learning
- Speech Analysis/Synthesis
- Speech Recognition/Understanding
- Spoken Dialog Systems
- Information Retrieval and Extraction
- Web Information Extraction/Mining
- Corpus Linguistics
- Multilingual/Cross-lingual Language Processing

Membership & Subscriptions

If you are interested in joining ACLCLP, please see appendix for further information.

Copyright

Cover

Calligraphy by Professor Ching-Chun Hsieh, founding president of ACLCLP
Text excerpted and compiled from ancient Chinese classics, dating back to 700 B.C.
This calligraphy honors the interaction and influence between text and language

International Journal of
Computational Linguistics and Chinese Language Processing
vol. 23, no. 1, June 2018

Contents

Computational Linguistics and Chinese Language Processing
Vol. 23, No. 1, June 2018, pp. 1-18

Sentiment Analysis on Social Network: Using Emoticon Characteristics for Twitter Polarity Classification

Chia-Ping Chen*, **Tzu-Hsuan Tseng*** **and Tzu-Hsuan Yang***

Abstract

In this paper, we describe a sentiment analysis system implemented for the semantic-evaluation task of message polarity classification for English on Twitter. Our system contains modules of data pre-processing, word embedding, and sentiment classification. In order to decrease the data complexity and increase the coverage of the word vector model for better learning, we perform a series of data pre-processing tasks, including emoticon normalization, specific suffix splitting, and hashtag segmentation. In word embedding, we utilize the pre-trained word vector provided by GloVe. We believe that emojis in tweets are important characteristics for Twitter sentiment classification, but most pre-trained sets of word vectors contain few or no emoji representations. Thus, we propose embedding emojis into the vector space by neural network models. We train the emoji vector with relevant words that contain descriptions and contexts of emojis. The models of long short-term memory (LSTM) and convolutional neural network (CNN) are used as our sentiment classifiers. The proposed emoji embedding is evaluated on the SemEval 2017 tasks. Using emoji embedding, we achieved recall rates of 0.652 with the LSTM classifier and 0.640 with the CNN classifier.

Keywords: Sentiment Analysis, Polarity Classification, Machine Learning, Neural Network, Word Embedding.

1. Introduction

There has been huge growth in the use of social networks, such as Twitter, in recent years. Many messages are created every day, including various topics, users' comments and views, or current emotions. Sentiment analysis, which predicts the polarity of a message, is one of the research directions on Twitter. A message on Twitter is called a tweet and is allowed to be 140 characters or less. Tweets are highly colloquial. Due to the length constraint, a tweet often

* Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan
E-mail: cpchen@cse.nsysu.edu.tw; fb74123698@gmail.com; kr60903@gmail.com

contains unofficial abbreviations, as well as emoticons and emojis. Figure 1 shows examples of tweets.

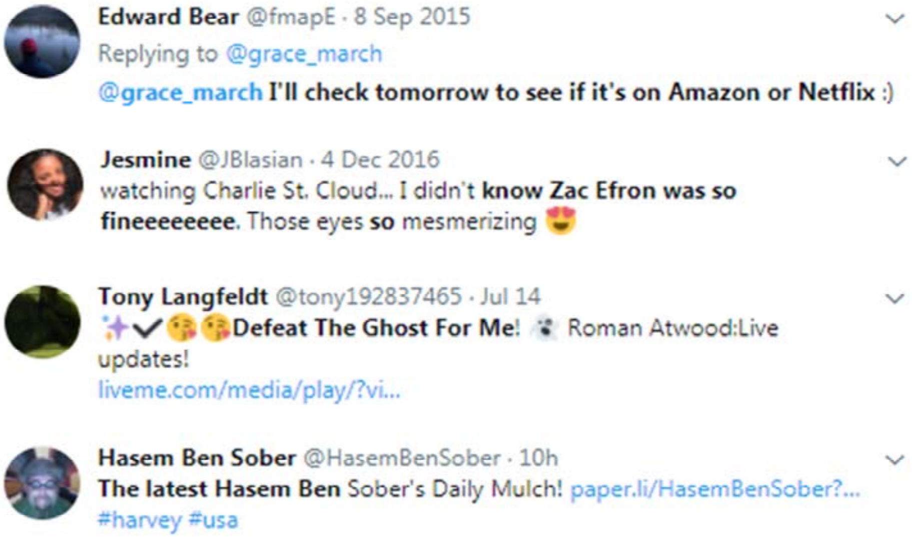


Figure 1. Examples of tweets. Tweets tend to have informal words and syntax.

In the above examples, we can see that emojis are used frequently in Tweets. Some emojis (like 😀) can be considered the natural evolution of emoticons, such as :-) and :D. In addition to facial expressions, emojis can be used for food, flags, animals, *etc*.

Unofficial abbreviations and emojis without corresponding word vectors in tweets can make the sentiment classification task difficult. In this work, we find sentiment features for these unorthodox tokens to get better results in sentiment classification.

Artificial neural networks for machine learning are mathematical models inspired by biological neural systems. Deep learning, which is neural network models based on deep neural networks, has been a very successful method and achieves state-of-the-art performance in many tasks, such as NIST handwritten digit recognition (LeCun, Bottou, Bengio, & Haffner, 1998) and ImageNet image classification (Krizhevsky, Sutskever, & Hinton, 2012). It performs well in natural language processing tasks, such as machine translation (Sutskever, Vinyals, & Le, 2014) and handwriting recognition (Graves *et al*., 2009).

For sentiment analysis, deep learning-based approaches have performed well in recent years. For example, convolution neural networks (CNN) with word embedding have been implemented for text classification (Kim, 2014), and they have achieved state-of-the-art results in SemEval 2015 (Severyn & Moschitti, 2015). SemEval 2017 Task 4 is sentiment analysis in Twitter, which is further divided into five subtasks: message polarity classification (Subtask A), topic-based message polarity classification (Subtasks B-C), and tweet quantification (Subtasks D-E).

Most of the participants in SemEval who adopt deep learning collect millions of tweets to train word-embedding models. The top system of SemEval 2017, which achieved 0.681 of average recall, used 100 million unlabeled tweets to pre-train word-embedding models (Cliché, 2017). In contrast, our goal in this work is to achieve sound performance without a large amount of external data.

In this paper, we describe our system for SemEval 2017 Task 4 (Subtask A) for message polarity classification (Rosenthal, Farra, & Nakov, 2017). Given a message, the system decides whether the message is of positive, negative, or neutral sentiment. We extend our previous work on SemEval 2017 (Yang, Tseng, & Chen, 2017). Our system consists of data pre-processing, word embedding, and classifiers. Data pre-processing includes normalization and hashtag segmentation. We consider the importance of emojis for sentiment analysis. In addition to using pre-trained word vectors, we train the emoji vector by the neural network. For the classifiers, we choose RNN-based and CNN models. We have achieved average recall rates of 0.652 with the LSTM-based classifier, and 0.640 with the CNN-based classifier.

Our contributions are described as follows.

- We propose neural network models for emoji embedding and investigate the effects of using emoji vectors in the classifiers. Through experiments, we find that emoji vectors can improve the accuracy of prediction for the positive and negative classes.
- Besides adding emoji vectors in the system, data pre-processing is critical to the improvement of the average recall rate from 0.610 to 0.652 with the LSTM classifier. Data pre-processing is important for Twitter sentiment analysis because textual data on Twitter is informal. In particular, the effect of hashtag segmentation is the most significant.

This paper is organized as follows. In Section 2, we describe our system, consisting of data pre-processing, word embedding, emoji embedding, and classifiers. In Section 3, we introduce data in experiments, network settings, and tools. In Section 4, we present the evaluation results, along with our comments. In Section 5, we conclude and discuss future works.

2. Related Work

There has been considerable research in the field of sentiment analysis. Past research mostly has focused on long text. Pang, Lee and Vaithyanathan (2002) analyzed the performance on movie reviews using machine learning algorithms and used star ratings as polarity signals in their training data. In recent years, there have been many research projects of sentiment analysis on social networks like Twitter. Go, Bhayani and Huang (2009) used distant learning to acquire more sentiment data. Their training data consisted of tweets with emoticons, which can be used as noisy labels. They constructed models with Naïve Bayes, Maximum Entropy, and Support Vector Machines (SVM), and they concluded that SVM outperforms the other models.

Deep learning has gained much attention in classification of Twitter text data, due to its huge success in speech recognition and computer vision. Among the top systems of SemEval, Severyn and Moschitti (2015) proposed a parameter initialization method for CNN. They used an unsupervised neural language model to initialize word embeddings that were fine-tuned by a distant supervised corpus. The pre-trained parameters were used to initialize the CNN model. Deriu, Gonzenbach, Uzdilli, Lucchi, De Luca and Jaggi (2016) utilized large amounts of data with distant supervision to train an ensemble of two-layer convolutional neural networks whose predictions were combined using a random forest classifier. Cliché (2017) used CNN models and bi-directional LSTM models. They pre-trained word embedding and fine-tuned it using distant supervision. They trained their models on Twitter data where embedding was fine-tuned again and finally combined several CNNs and LSTMs to get better performance.

Emojis are an important feature in tweets. Many studies have analyzed and trained emojis, such as Zhao, Dong, Wu and Xu (2012) and Barbieri, Ronzano and Saggion (2016). Zhao *et al*. (2012) built a system, which was the first system for sentiment analysis of Chinese tweets in Weibo. They mapped 95 emojis into four categories of sentiment. Their system employs the emojis for the generation of sentiment labels for tweets and builds an incremental learning Naïve Bayes classifier for the categorization of four types of sentiments. Barbieri *et al*. (2016) studied Twitter emojis with embedding models. They retrieved ten million tweets posted by USA users, and they made vector models of both words and emojis using several skip-gram word-embedding models.

3. Method

The system we implement for sentiment classification is shown in Figure 2. In data pre-processing, we normalize data sets to decrease the data complexity and increase the coverage of the word vector inventory. In word embedding, we utilize pre-trained word vectors provided by GloVe (Pennington, Socher, & Manning, 2014) and we train emoji embedding by neural networks. The models of LSTM and CNN are used as our sentiment classifiers.

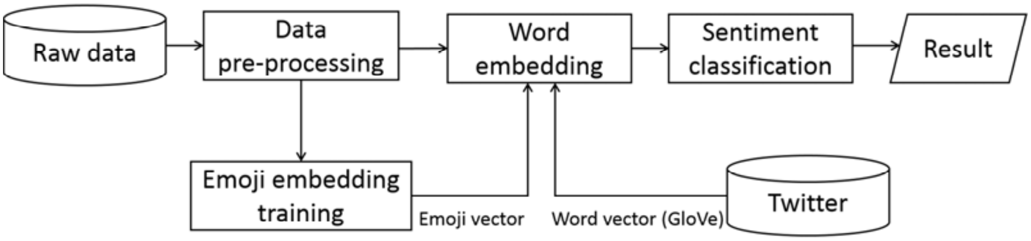


Figure 2. Our Sentiment System. We propose to add embedded emoji vectors for better sentiment classification.

3.1 Data Pre-processing

All the data used for training the emoji embedding and for training the sentiment classification models undergo a series of data pre-processing. First, we use a tokenizer to split a tweet into words, emoticons, and punctuation marks. Happytokenizer[1] is the tokenizer we use for text processing. Then, we replace URLs and USERs with normalization patterns <URL> and <USER>, respectively. All uppercase letters are converted into lowercase letters. The above pre-processing is called basic pre-processing. Next, we perform further data pre-processing based on basic pre-processing. The further data pre-processing is described as follows.

3.1.1 Emoticon Normalization

Tweets often contain a variety of emoticons, and some emoticons do not correspond to any pre-trained word vector. To reduce complexity, we normalize similar emoticons to the same token, as described in Table 1.

Table 1. Examples of emoticon normalization. We normalize similar emoticons into four categories, which are <smile>, <sadface>, <neutralface>, and <heart>, respectively.

Emoticon	Normalization
:), (:, :-), (-:, :D, :-D, ;), =), (=, =D	<smile>
:(,):, :'(,)':, =(,)=, :-(,)-:	<sadface>
:|, |:, =|, |:	<neutralface>
<3	<heart>

For example, in the case of <neutralface> category, we find :|, |:, =|, |:, and then replace them by the token <neutralface>. Thus, the emoticons are replaced by four normalized categories.

[1] http://sentiment.christopherpotts.net/tokenizing.html

3.1.2 Specific Suffix Splitting

There are many specific suffixes in English words, such as children's and Amazon's. These words may not have any corresponding word vectors and their presence increases the vocabulary size. So, we split the specific suffixes, including 's, n't, 'll, 're, 've, 'd and 'm, to decrease vocabulary. Moreover, the words resulted from splittingy do have corresponding word vectors (Nabil, Atyia, & Aly, 2016).

3.1.3 Hashtag Segmentation

Hashtag often is composed of multiple words and includes emotional words. We try two ways to split hashtags. Table 2 shows the examples of hashtag segmentation.

♦ Maximum Matching Segmentation

We use the vocabulary of GloVe as our dictionary containing approximately 570,000 words, and define a regular expression for numbers and punctuation. From the beginning of a hashtag, we split it according to the dictionary as much as possible until segmentation is finished.

♦ Unigram-based Segmentation

We train a unigram model with 0.6M tweets obtained from Twitter API. We do statistics of different words in these tweets and remove Except for the letter “a” and the letter “i”, single letters letters are removed from the unigram dictionary to avoid over-segmentation. All results of hashtag segmentation will be split according to the dictionary, so a hashtag may have multiple results. Then, we calculate sum of log probability of each word from each result as segmentation score. Finally, we take the highest score as the final segmentation result.

Table 2. Examples of hashtag segmentation. A hashtag is converted to a word sequence.

Hashtag	Maximum matching	Unigram-based
#windows10fail	windows 10 fail	windows 10 fail
#sportshalloweencostume	sports Halloween costume	sports Halloween costume
#thisisnotajoketweet	thisis notajoke tweet	this is not a joke tweet

3.2 Embedding

Since training word embedding requires a lot of time, we use the pre-trained word vector provided by GloVe. Nevertheless, many pre-trained sets of word vectors contain few or no emoji representations. Therefore, Barbieri *et al*. built skip-gram word embedding models by mapping both words and emojis in the same vector space (Barbieri *et al*., 2016). Note that we only consider emojis and do not include emoticons because emoticons already are normalized to the normalization tokens during the data pre-processing phase. We train emoji vectors by neural networks. Figure 3 shows the model architecture. In Figure 3, ***U*** is the weight matrix from the input layer to the hidden layer and ***V*** is the weight matrix from the hidden layer to the output layer. When the embedding training is finished, the weight matrix from the input layer to the hidden layer ***U*** consists of the emoji vectors. We use pairs of an emoji and its relevant words, including the descriptive words and the contextual words, as training examples. The steps are described as follows.

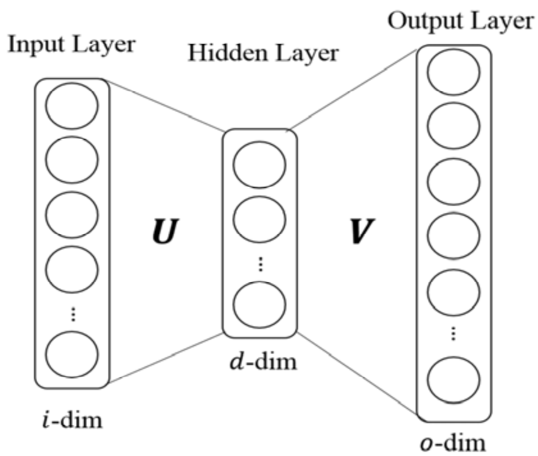


Figure 3. The model of emoji embedding. Similar to word embedding, an emoji vector is trained to predict neighboring word vectors.

3.2.1 Description Words

We crawled emojis and their descriptions from Unicode emoji standard,[2] resulting in 9,244 description words for 2,623 emojis. Every training example consists of an emoji and a sequence of words $w_1, w_2, \ldots, w_n$ describing that emoji. We tried two methods of producing the training target, described as follows. Table 3 shows examples of emojis and their descriptions.

Table 3. Examples of emojis and their description.

Emoji	Description
😃	Grinning face
😁	Beaming face with smiling eyes
🏃🏽‍♀️	Woman running: medium skin tone
💦	Sweat droplets
🐱	Cat face

[2] http://www.unicode.org/emoji/charts/full-emoji-list.html

♦ Sum of the vectors of the description words

We take the sum of the individual vectors of description words as a training target, where the word vector can be found in GloVe. The description words $w_{1:n}$ correspond to pre-trained word vectors $\boldsymbol{v}_{1:n}$ where $\boldsymbol{v}_i$ is a d-dimensional word vector. The training target is $= \sum_{i=1}^{n} \boldsymbol{v}_i$. In this way, the number of training samples is 2,623.

♦ Description word splitting

We divide the description words into n training examples. For example, the description of 😃 is grinning face, so the training examples are (😃, grinning) and (😃, face). These description words have corresponding pre-trained word vectors. In this way, the number of training examples is 9,244.

The size of the input layer is equal to the number of different emojis, as emojis are represented by one-hot vectors. The size of the output layer is equal to the size of word vector, which is 100.

3.2.2 Contextual Words

We selected tweets with emojis from the aforementioned 0.6M tweets, resulting in a set of approximately 50K tweets. A tweet is $w_1, w_2, \ldots, e, \ldots, w_n$, where e is an emoji. The network input is emoji e as a one-hot vector, and the network output targets are the contextual words of e, which is $w_1, w_2, \ldots, w_n$. Training examples are $(e, w_1), (e, w_2), \ldots, (e, w_n)$. We collected about 1.7M training samples.

For the output target, we tried two kinds of representations. A sparse target representation uses a one-hot vector, while a distributed target representation uses a word vector.

3.3 Sentiment Classification

3.3.1 LSTM Classifier

Figure 4 shows the architecture of our RNN-based classifier, which contains an input layer, embedding layer, hidden layer, and soft-max layer.

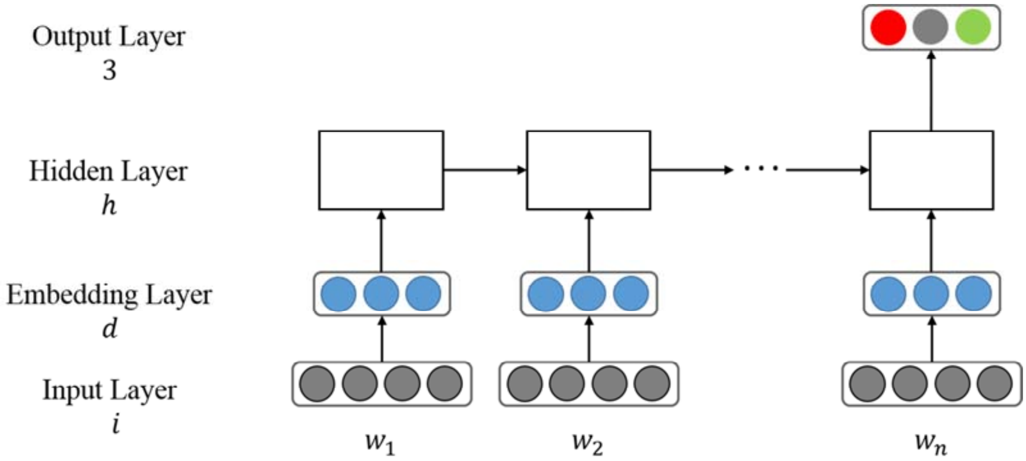


Figure 4. The architecture of the implemented classifier based on recurrent neural network with long short-term memory (LSTM) cells in the hidden layer.

In the input layer, each tweet is treated as a sequence of words and each word is input into the model at every time step. In the embedding layer, each word is converted to a word vector, where word vectors are stored in an embedding matrix provided by GloVe. In the hidden layer, we use LSTM memory cells (Hochreiter & Schmidhuber, 1997) for the long-range dependency. Different from the original recurrent unit, the LSTM cell contains gates to control states. The hidden states of the first word to the second to last word in a tweet connect to the hidden state of the next word. Only the hidden state of the last word connects to the next (output) layer. In the soft-max layer, output values are processed by soft-max function to get probabilities for classification.

3.3.2 CNN Classifier

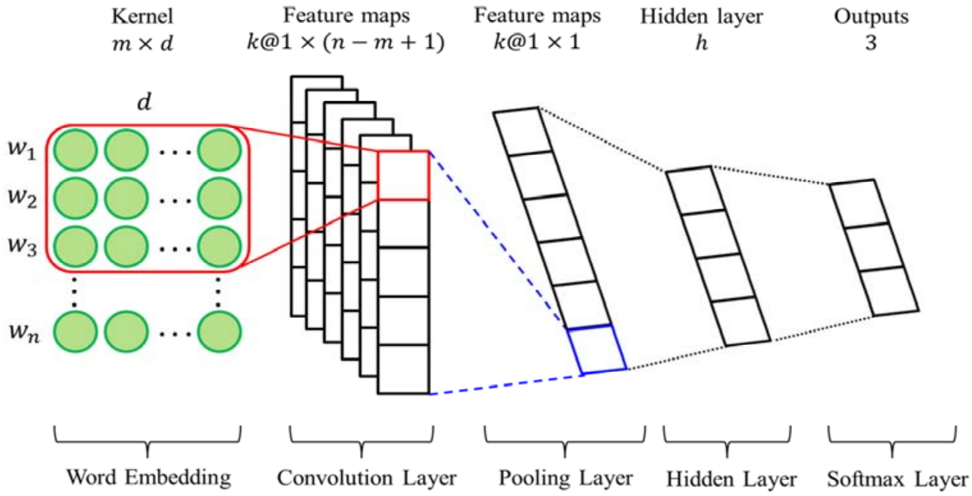


Figure 5. The architecture of the implemented classifier based on convolutional neural network (CNN).

The CNN model we use for classification is the architecture used by Kim (2014). Figure 5 shows the CNN architecture, which consists of a convolutional layer, max-pooling layer, hidden layer, and soft-max layer.

The model input is a tweet, consisting of sequence of words $w_{1:n} \in \mathcal{R}^V$, where V is the vocabulary size. In order to fix the length of the tweet, we pad input text with zeros into length n. Each tweet $w_{1:n}$ is represented by the corresponding word vector $x_{1:n}$, where x_i is the d-dimension word vector of the i-th word. Input words are embedded into dense representation through word embedding and fed into the convolutional layer. A word without an embedding vector is represented by a zero vector. After word embedding, an input tweet is mapped to an input matrix $S \in \mathcal{R}^{n \times d}$.

In the convolutional layer, kernel $K \in \mathcal{R}^{d \times m}$ slides over the input matrix with stride $s = 1$ and creates features c_i.

$$c_i = f(K * S_{i:i+m-1} + b_{conv}) \quad (1)$$

where b_{conv} is the bias at the convolutional layer, * denotes the convolution operation, and f is a nonlinear function. The feature map $y_{conv} \in \mathcal{R}^{n-m+1}$ is created by

$$y_{conv} = [c_1, c_2, \dots, c_{n-m+1}] \quad (2)$$

We use k kernels to create k feature maps, which are denoted by $Y_{conv} \in \mathcal{R}^{k \times (n-m+1)}$. Then, we apply the max-pooling operation over each feature map in order to capture important information, *i.e.*

$$y_{pool,i} = \max_j Y_{conv,i,j} \quad (3)$$

where $y_{pool,i} \in \mathcal{R}^k$ is the output after the max-pooling operation.

After the max-pooling layer, we use dropout to drop some activations for regularization randomly in order to prevent the model from over-fitting. Finally, we use a fully connected layer of size h followed by a dense layer with soft-max function for classification.

4. Results

4.1 Data

We used the SemEval 2017 data provided by the task organizers. These data are tweets in Twitter, which are labelled with three types of sentiment: positive, neutral, and negative. The training data were the tweets from SemEval 2013 to SemEval 2016, excluding SemEval 2016 testing data. The development data were the tweets from SemEval 2016 testing data. The test data were the tweets provided by the organizer of SemEval 2017. Table 4 summarizes the statistics of the data.

Table 4. Statistics of SemEval 2017: the number of tweets in datasets.

Data	Pos.	Neu.	Neg.	Total
Train	12,844	12,249	4,609	29,702
Dev	7,059	10,342	3,231	20,632
Test	2,375	5,937	3,972	12,284

4.2 Settings

We used the pre-trained word vectors provided by GloVe, which are trained with Twitter data. The dimension of word vectors can be 50, 100, or 200. We evaluated these dimensions with the SemEval 2016 dataset. The 100 and 200-dimension word vectors achieved better results, so we used 100-dimension word vectors for the SemEval 2017 tasks. We noticed that the performance is not very sensitive to the hyper-parameter of word vector size and the number of hidden layer units. For the CNN model, the number of filters k was 50. The kernel size was 3×100 with stride $s = 1$ over the input matrix. Max-pooling was applied over each feature map. After pooling, we dropped activations randomly with the probability of $p = 0.2$ and fed to the hidden layer with size $h = 20$. The hyperbolic tangent function was used as the activation function after convolution and pooling. For the LSTM model, input size i was equal to the size of word list and the size of hidden h was 50. We dropped input units for input gates and recurrent connections with the same probability of $p = 0.2$.

Next, we explain the settings of emoji embedding training. The size of input layer was equal to the number of emojis. For training with descriptive words, the size of the input layer was 2,623, the size of the hidden layer was 100, and the size of the output layer was 100. Both hidden-layer outputs and output-layer outputs went through hyperbolic tangent function, and the loss is the mean square error.

For training with contextual words, the size of the input layer was 1,023 and the size of the hidden layer was 100. The size of the output layer was 42,670 if the target was represented by one-hot vector and was 100 if the target was represented by word vector. Output values went through the soft-max function. Both hidden-layer outputs and output-layer outputs went through the hyperbolic tangent function, and the loss is the mean square error.

All the models we used in our experiments were implemented using Keras[3] with Tensorflow[4] backend.

4.3 Baseline

We participated in SemEval 2017 task 4, which is sentiment classification in Twitter (Yang *et al*., 2017). There are three evaluation measures in the task, which are average recall of three classes, average F-measure of positive and negative classes, and accuracy. The organizer chose average recall as the primary measure because it is more robust to class imbalance (Rosenthal *et al*., 2017), so we focus on this performance measure. The results of each setting are obtained from an average of five runs of experiments. In each run, we trained our models with the same usage of data sets, instead of the cross-validation or leave-one-out scheme.

Table 5 shows our results of SemEval 2017. We interpolated the LSTM and CNN models to get the interpolated model for the final submission, which achieved 0.618 for average recall. Also, we list the evaluation of LSTM and CNN model. We will take this performance as our baseline.

Table 5. Results of baseline. Interpolation-baseline is the result of participating in SemEval 2017. It is an interpolation of an LSTM model and a CNN model. The result of LSTM-baseline and the result of CNN-baseline are obtained from an average of five runs of experiments.

Model	Avg. Recall	Avg. F1	Accuracy
LSTM-baseline	0.610	0.575	0.615
CNN-baseline	0.584	0.548	0.583
Interpolation-baseline	0.618	0.587	0.616

[3] https://keras.io

[4] https://www.tensorflow.org

4.4 Comparison of Data Pre-processing

In this part, we show the classification results using different data pre-processing. We summarize different statistics of pre-processed data, including vocabulary size, the number of tokens in data, and coverage on word vector. Here, coverage on word vector is the proportion of tokens with found word vectors. The word list we used was extracted from the words in the training data, and it is equal to vocabulary size. If the word vector of a word in the word list could not be found in GloVe, we used the zero vector for the word. Words in the testing data but not in the word list were removed. Each LSTM or CNN model was trained with 50 epochs. Table 6 shows the results of data pre-processing.

Table 6. Statistics of pre-processed data and results. In each pre-processing stage, we base on basic pre-processing. "emoticon" means emoticon normalization. "suffix" means specific suffix splitting. "hashtag*" means hashtag segmentation, where hashtag1 is maximum matching segmentation and hashstag2 is unigram-based segmentation.

Pre-processing	# Vocab.	# Tokens		Coverage		Avg. Recall	
		Train	Test	Train	Test	LSTM	CNN
Basic	38,353	691,261	218,821	0.914	0.886	0.610	0.584
Basic + emoticon	38,316	686,565	217,321	0.916	0.886	0.615	0.584
Basic + suffix	37,506	700,736	222,435	0.928	0.904	0.614	0.584
Basic + hashtag1	36,429	695,654	224,562	0.924	0.932	0.622	0.590
Basic + hashtag2	34,327	700,954	232,322	0.924	0.933	0.616	0.593
Basic + emoticon + suffix + hashtag1	35,542	700,410	226,621	0.940	0.948	0.625	0.594

It can be seen that the vocabulary size decreased and the coverage on word vector increased after data pre-processing. For the LSTM model, average recall of all pre-processed data was higher than the basic pre-processed data, especially data with hashtag segmentation. For the CNN model, average recall also was improved by data with hashtag segmentation. Although hashtag with unigram-based segmentation can attain better results, its average recall was lower than hashtag with maximum matching segmentation in the LSTM model. We think that the segmentation dictionary we used in maximum matching was the vocabulary of GloVe, so most words after hashtag segmentation had corresponding vectors. We combined all data pre-processing steps finally and achieved 0.625 for the LSTM model and 0.594 for the CNN model.

4.5 Evaluation of Emoji Vector

In this part, we explore the effect of adding the emoji vector. In order to have more corresponding vectors in testing data, the word list we used in this experiment was the vocabulary of GloVe. Words in the training and testing data but not in the word list were removed. If emoji vectors were added, we added the emoji to the word list.

In order to prevent the model from over-training, we used an early stopping mechanism. We added the development dataset as validation data during model training. If there was no improvement in accuracy of validation data, the model stopped training. With early stopping, each model was trained about 5-10 epochs.

The results of emoji embedding are shown in Table 7. 'No emoji' means that the emoji vector was not added. 'desc_sum' means that the sum of the description word vectors of emoji is the emoji vector. For training the emoji vector with description words, the sum of the individual word vectors of description words as a training target and dividing description words into multiple training targets are denoted as 'desc_sum_nn' and 'desc_sp_nn,' respectively. For training the emoji vector with contextual words, the training target using one-hot vector and word vector are denoted by 'skip gram_1H' and 'skip gram_vec,' respectively.

Table 7. Results of emoji embedding. Note that the refined system with 0.652 would have ranked 7th in SemEval 2017.

Pre-processing	Basic						All					
Model	LSTM			CNN			LSTM			CNN		
Early stopping	$AvgR$	F_1^{PN}	Acc.	$AvgR$	F_1^{PN}	Acc.	$AvgR$	F_1^{PN}	Acc.	$AvgR$	F_1^{PN}	Acc.
No emoji	0.634	0.601	0.627	0.624	0.594	0.618	0.651	0.630	0.639	0.629	0.599	0.612
desc_sum	0.635	0.606	0.632	0.628	0.601	0.620	0.651	0.626	0.638	0.630	0.605	0.622
desc_sum_nn	0.639	0.611	0.633	0.626	0.592	0.608	0.645	0.615	0.628	0.631	0.601	0.617
desc_sp_nn	0.638	0.607	0.623	0.626	0.596	0.615	0.652	0.628	0.632	0.638	0.618	0.623
skip gram_1H	0.639	0.613	0.633	0.622	0.593	0.617	0.642	0.614	0.633	0.640	0.618	0.620
skip gram_vec	0.639	0.616	0.635	0.620	0.586	0.610	0.646	0.619	0.626	0.633	0.604	0.614

By adding the emoji vector in systems with basic pre-processing, the average recall of the two models was mostly better than with no emoji. With all of the pre-processing, there was no significant improvement in the LSTM model. In CNN models with all of the pre-processing, the performance of adding emoji vectors was still better than without the emoji vector.

We know that only tweets in test data have emoji, and there were 731 tweets with emoji, which made about 5.9% of the test data. Furthermore, models did not learn emoji characteristics directly during training because there were no tweets with emoji in the training data. These are possible reasons there was no significant improvement in some cases.

In order to more clearly observe the effect of adding emoji vectors for model classification, we only evaluated the test data with emojis in previous LSTM and CNN models. Table 8 shows the statistics of tweets with emoji. Table 9 shows the evaluation of tweets with emoji.

Table 8. Statistics of tweets with emoji. In our data set, only tweets in test data had emoji.

Tweet with emoji	Pos.	Neu.	Neg.	Total
Test	310	248	173	731

Table 9. Evaluation of tweets with emoji.

Pre-processing	Basic						All					
Model	LSTM			CNN			LSTM			CNN		
Early stopping	Avg$R$	F_1^{PN}	Acc.	Avg$R$	F_1^{PN}	Acc.	Avg$R$	F_1^{PN}	Acc.	Avg$R$	F_1^{PN}	Acc.
No emoji	0.621	0.665	0.638	0.621	0.661	0.630	0.644	0.696	0.656	0.624	0.679	0.637
desc_sum	0.636	0.682	0.639	0.611	0.651	0.605	0.648	0.701	0.651	0.615	0.661	0.617
desc_sum_nn	0.629	0.665	0.637	0.601	0.640	0.616	0.633	0.679	0.650	0.617	0.669	0.635
desc_sp_nn	0.599	0.639	0.614	0.606	0.647	0.617	0.630	0.682	0.645	0.624	0.683	0.635
skip gram_1H	0.632	0.666	0.636	0.604	0.641	0.612	0.638	0.685	0.648	0.634	0.692	0.643
skip gram_vec	0.611	0.637	0.624	0.591	0.622	0.608	0.627	0.670	0.649	0.619	0.665	0.636

The results show that the effect of adding emoji is not obvious. From our observation on the performance of the three classes, the addition of emoji vectors decreases the prediction of neutral class dramatically, but increases the prediction of positive and negative classes. Besides, we found that emoji vectors are more similar to each other than to the word vectors in the embedding space. Thus, they can contribute supplementary information for sentiment analysis.

5. Conclusion

We implemented our sentiment analysis system for sentiment analysis of Twitter data organized in SemEval 2017. This system consists of data pre-processing, word and emoji embedding, and classifier. From our observation, the data complexity decreases after data pre-processing with improved performance on classification, especially with hashtag segmentation. We found that adding emoji vectors can improve the performance on classification, especially for CNN models, and model training with an early stopping mechanism can prevent the model from over-training.

In data pre-processing, we process data with basic pre-processing and all pre-processing, including emoticon normalization, specific suffix splitting, and hashtag segmentation. In word embedding, we train emoji embedding with the descriptive words or the contextual words of emojis. For our models, we set the vocabulary of GloVe as the word list of models instead of the vocabulary in training data and validation data. In addition, we used an early stopping mechanism to train our models. Our system achieved 0.652 for LSTM model and 0.640 for CNN model, which would have ranked 7th in SemEval 2017.

Regarding future works, we hope to get closer in performance to the leaders on the task leader-board. As mentioned in Section 4.5, the models did not learn emoji characteristics directly during training. Thus, we want to collect tweets with emojis for training and do further evaluation on our models. We also will try the fine-tuned word vector and make it suitable for sentiment classification.

Reference

Barbieri, F., Ronzano, F., & Saggion, H. (2016). What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis. In *Proceedings of the LREC* 2016.

Black, E., Abney, S., Flickenger, D., Gdaniec, C., Grishman, R., Harrison, P., ... Strzalkowski, T. (1991). A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars. In *Proceedings of the Workshop on Speech and Natural language (HLT '91)*, 306-311. doi: 10.3115/112405.112467

Cliché, M. (2017). BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017)*. doi: 10.18653/v1/S17-2094

Deriu, J., Gonzenbach, M., Uzdilli, F., Lucchi, A., De Luca, V., & Jaggi, M. (2016). SwissCheese at SemEval-2016 Task 4: Sentiment Classification Using an Ensemble of Convolutional Neural Networks with Distant Supervision. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, 1124-1128. doi: 10.18653/v1/S16-1173

Go, A., Bhayani, R., & Huang, L. (2009). *Twitter Sentiment Classification using Distant Supervision*. CS224N Project Report, Stanford, 1(12).

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009). A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE transactions on pattern analysis and machine intelligence*, *31*(5), 855-868. doi: 10.1109/TPAMI.2008.137

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in neural information processing systems (NIPS'12)*, 1097-1105.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based Learning Applied to Document Recognition. In *Proceedings of the IEEE*, *86*(11), 2278-2324. doi: 10.1109/5.726791

Nabil, M., Atyia, A., & Aly, M. (2016). CUFE at SemEval-2016 Task 4: A Gated Recurrent Model for Sentiment Classification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, 52-57. doi: 10.18653/v1/S16-1005

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment Classification using Machine Learning Techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing (EMNLP '02)*, 10, 79-86. doi: 10.3115/1118693.1118704

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. doi: 10.3115/v1/D14-1162

Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017)*, 502-518.

Severyn, A., & Moschitti, A. (2015). UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 464-469. doi: 10.18653/v1/S15-2079

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In *Advances in neural information processing systems (NIPS'14)*, 3104-3112.

Yang, T. H., Tseng, T. H., & Chen, C. P. (2017). deepSA at SemEval-2017 Task 4: Interpolated Deep Neural Networks for Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017)*, 616-620. doi: 10.18653/v1/S17-2101

Zhao, J., Dong, L., Wu, J., & Xu, K. (2012). Moodlens: an Emoticon-based Sentiment Analysis System for Chinese Tweets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12),* 1528-1531. doi: 10.1145/2339530.2339772

Computational Linguistics and Chinese Language Processing
Vol. 23, No. 1, June 2018, pp. 19-34

長短期記憶模型之忘記閘提取語意流暢度之架構以自閉症小孩說故事為例

A Lexical Coherence Representation Computational Framework using LSTM Forget Gate For Autism Recognition

劉于碩*、陳金博*、高淑芬+、李祈均*

Yu-Shuo Liu, Chin-Po Chen,

Susan Shur-Fen Gau and Chi-Chun Lee

摘要

泛自閉症研究指出，由於口語表達能力的遲緩，跟典型孩童相比，自閉症孩童較無法敘說一個流暢的故事，因此在診斷自閉症孩童時，衡量其口語表達流暢度以成為一個診斷的重要指標。然而流暢度的評量，不是需要費時的人工標註，就是得用專家專門設計出的特徵來當指標，因此，本篇研究提出一種自動化直接資料導向的流暢度特徵學習架構，利用長短期記憶模型的遺忘閘導出語意流暢度的特徵，同時我們也利用自閉症觀察診斷量表中的評分細項來測試我們的流暢度特徵，結果上，利用我們提出的語意流暢度特徵來辨識自閉症小孩與典型小孩的任務上能夠達到 92%的高準確率，對照傳統上使用語法、語詞使用頻率、潛在語義模型分析等地模型有顯著的提升。

* 國立清華大學電機工程學系
Department of Electrical Engineering, National Tsing Hua University
E-mail: drugstore950337@gmail.com; cclee@ee.nthu.edu.tw
+ 國立台灣大學醫學院心理系
Department of Psychiatry, National Taiwan University Hospital and College of Medicine
E-mail: gaushufe@ntu.edu.tw

這篇論文也進一步隨機打亂字序及句子順序，使典型小孩說故事內容變得不流暢的方式，來驗證我們提出的流暢度的意義，降維並將資料樣本可視化分析後證明我們的提取的特徵含有流暢度的概念。

Abstract

Autistic children are less able to tell a fluent story than typical children, so measuring verbal fluency becomes an important indicator when diagnosing autistic children. Fluency assessment, however, needs time-consuming manual tagging, or using expert specially designed characteristics as indicators, therefore, this study proposes a coherence representation learned by directly data-driven architecture, using the forget gate of long short-term memory model to export lexical coherence representation, at the same time, we also use the ADOS coding related to the evaluation of narration to test our proposed representation. Our proposed lexical coherence representation performs high accuracy of 92% on the task of identifying children with autism from typically development. Comparing with the traditional measurement of grammar, word frequency, and latent semantic analysis model, there is a significant improvement.

This paper also further randomly shuffles the word order and sentence order, making the typical child's story content become disfluent. By visualizing the data samples after dimension reduction, we further observe the distribution of these fluent, disfluent, and those artificially disfluent data samples. We found the artificially disfluent typical samples would move closer to disfluent autistic samples which prove that our extracted features contain the concept of coherency.

關鍵詞：人類行為分析、流暢度、長短期記憶模型、自閉症、說故事

Keywords: Behavioral Signal Processing, Lexical Coherence Representation, LSTM, Autism Spectrum Disorder, Story-telling.

1. 緒論 (Introduction)

自閉症是一種普遍且又病症複雜的精神症候群，不同於一般較單純的精神疾病，每個自閉症患者之間的症狀異質性非常的高，此一特性也增加了診斷的困難性，因此許多研究致力於設計出各種評量的標準與特徵來做為人工自閉症診斷或者自動化自閉症診斷的標準，此類研究對於自閉症的即早治療與縮短診斷時程有極大幫助，也因此在人類行為訊號處理的領域裡越發蓬勃發展。過去許多研究已針對自閉症與典型孩童的各種不同模態行為進行過分析，例如利用臉部不同的表情反應特徵來辨識自閉症的研究(Liu, Li & Yi, 2016)，以及利用孩童之間不同的語音表徵(Marchi *et al.*, 2015; Bone *et al.*, 2014)，諸如聲調或發音咬字等等，也有利用孩童敘事的文字內容去分析語法和用字作為分析自閉症辨

識特徵的研究(Regneri & King, 2016; Chorianopoulou *et al*., 2017)。我們在此篇論文中提出一種資料導向的語意流暢度特徵提取架構，這個架構能夠利用自閉症以及典型孩童說故事的資料辨識出說故事的資料來自於自閉症孩童或典型孩童。

過去對於自閉症孩童說故事的分析時常使用許多不同的關鍵字特徵或是主題特徵，來觀察自閉症與典型孩童的差異(Rouhizadeh, Prud'Hommeaux, Roark & Van Santen, 2013; Rouhizadeh, Sproat, & Van Santen, 2015)。許多研究更指出相對於典型孩童，自閉症孩童較無法說出一個完整的語法流暢的故事(Losh & Gordon, 2014)，或是自然地描述故事發展內容的因果關係。分析顯示，自閉症小孩描述故事時，傾向於把故事中的一個個元素、事件單獨描述，缺乏流暢的串接或敘述各個情節之間的因果關係(Capps, Losh, & Thurber, 2000; Tager-Flusberg, 1995)。而語法表達的不流暢現象也已有研究觀察證實(Diehl, Bennetto & Young, 2006)。而在一篇 ACL2016 的研究中更利用了許多關鍵字性質標註表達整個文章的劇情走向與分佈，藉由這個導出的劇情發展狀況來當作衡量孩童說故事的流暢性。然而，以上方法皆須仰賴經過訓練的專業人士標註，標註十分費時，且會受限於只能在被設計的情況下使用。

因此，我們的研究使用資料導向的方法，以文字向量當作輸入，輸進長短期記憶模型，使用期訓練好的遺忘閘作為參數，讓機器自己學出表達資料中含有流暢度的時序性語法特徵。實驗中，我們比較以各不同的特徵來訓練針對說故事資料庫的自閉症辨識模型的準確率，而我們所提出的語意流暢度特徵所訓練的模型在實驗中達到 0.92 的準確率，高過於由知名的流暢度特徵，流暢矩陣(Graesser, McNamara, Louwerse & Cai, 2004; McNamara, Graesser, McCarthy & Cai, 2014)所訓練出來的模型。再者，啟發於最新關於理解神經網路黑盒子的研究(Li, Monroe & Jurafsky, 2016; Koh & Liang, 2017)，我們觀察改變原始資料的性質後的語意流暢特徵分佈的改變，來試著理解推敲神經網路這個黑盒子運作，並驗證我們導出的表徵的意義。實驗中，我們將典型孩童所敘述故事的字序和句序分別打亂，來模擬不同程度的不流暢，結果驚訝的發現，在以我們導出的流暢度表徵空間裡，被打亂後的不流暢典型小孩說故事資料分佈，會往無法說出流暢故事的自閉症小孩說出的故事資料分佈靠近，此結果驗證了我們導出的表徵的確含有流暢度的資訊。此篇論文內容安排如下: 第二節為資料庫以及架構介紹，第三節為實驗結果分析與呈現，最後第四節為結論。

2. 研究方法 (Research Methodology)

2.1 資料庫 (Database)

這篇論文中，我們使用兩個資料庫，第一個是主要用來驗證我們所提取出的流暢度特徵辨識力的自閉症與典型孩童說故事資料庫，第二個則是從收錄了許多兒童讀物的童話故事網站，所爬下來的四種童話故事資料庫，第二個資料庫是用來預訓練使用。

2.1.1 資料庫一:自閉症與典型孩童說故事資料庫 (Database I)

資料庫一的蒐集方式是由專業培訓過的研究人員引導式的帶領自閉症孩童完成的，使用的故事書名為瘋狂星期二，故事書本身是一本繪本，書中每頁內容，都是一張圖片，與一句短句概要這頁的內容，而小孩子會根據圖片內容描述每一頁的劇情，完成整篇故事的論述。

而說故事的這個情境，便是標準的自閉症診斷觀察量表(Lord, Rutter, Dilavore & Risi, 2008)中的一個單元，此為一國際公認的臨床自閉症診斷量表，我們是與台大兒童醫院共同合作收集。這個量表是設計來觀測孩童的社交互動行為，因此設計了許多行為單元，來觀測並評量受測者的行為表現，單元包刮建構式作業、假扮遊戲、共同互動式遊戲、示範作業、圖片描述、看圖說故事等等。而在我們所使用的看圖說故事單元中，施測者會先引導一段引文，而孩童則會被要求根據後半部的繪本內容，完成後半部的故事，我們用相同的流程收集了典型孩童部分的說故事資料，逐字稿的內容舉例如:

“星期二的晚上，有一群青蛙在池塘上，他們突然乘著荷葉飛了起來，他們飛到了鄰近的小鎮，接下來換你說故事……”

在這個資料庫中我們總共收集了 67 位孩童的資料總共約有 28446 字的資料庫，每篇故事平均約有 424 字，67 位孩童中包括 31 位自閉症孩童(ASD)以及 36 位典型孩童(TD)。表 1 為資料庫相關敘述。

2.1.2 資料庫二:童話故事資料庫 (Database II)

資料庫二是從線上收藏不同種類童話故事的網站所收集而來，我們隨機選了四種篇幅長短約莫類似的故事種類，來作為我們預訓練的資料庫，這四種故事分別為俗諺故事、床邊故事、當代童話故事以及謎語故事。每種故事，我們都隨機選了 24 篇故事作為我們的預訓練資料集，總共 96 篇故事、59212 字，表 2 為相關的描述。

2.2 語意流暢度表徵 (Lexical Coherence Representation)

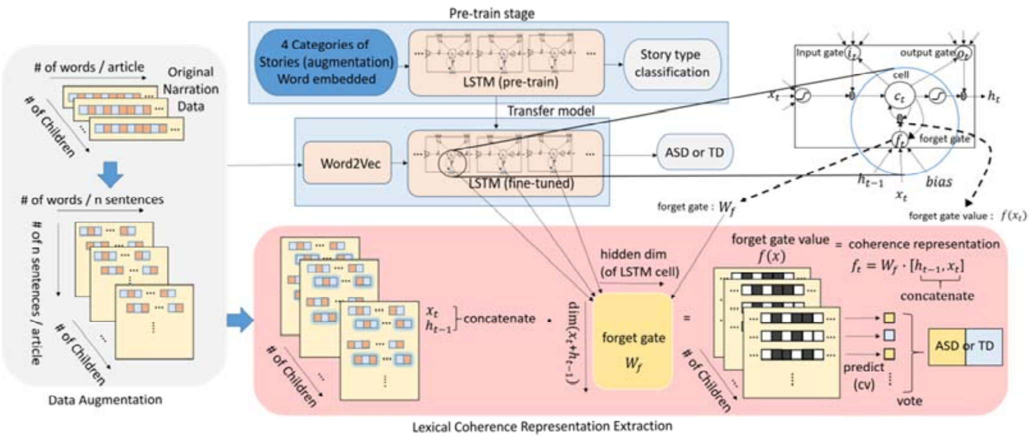


圖 1. 語意流暢度表徵提取實驗架構圖
[Figure 1. Architecture of lexical coherence representation]

圖 1 呈現了用來提取語意流暢度特徵的架構流程，流程包含了資料擴增、中文文字向量訓練模型、資料庫二之長短期記憶模型預訓練、資料庫一之長短期記憶模型適性訓練以及最終由遺忘閘提取語意流暢度特徵。

2.2.1 資料擴增 (Data Augmentation)

研究證實，透過資料擴增的方式能夠使的神經網絡看過更廣泛的資料分佈，進而有助於穩定神經網絡訓練的整體偏差率，在神經網絡翻譯與語音辨識的研究中都已證實這個方法的幫助性(Fadaee, Bisazza & Monz, 2017; Hannun *et al*., 2014)。在這篇文章中,我們使用一個簡單的滑動視窗方法來執行資料增加兩個資料集。我們用 n 個句子作為一筆資料，並使用與整個文檔相同標籤的樣本，我們每次移動一個句子來生成另一個 n 句的資料樣本。實驗也比較了由各種不同個句子數組成的資料樣本的結果，分別試驗了 n 為 1，3，5，7，9 的情況。

2.2.2 中文文字斷詞 (Chinese Word Segmentation)

近年來的文字相關研究中，都使用文字向量空間作為普遍公認的語意表徵，但不同於英文，中文字在書寫上，並沒有將詞語詞之間隔開，因此在進入文字向量的訓練之前，必須先經過斷詞的程式，這裡我們使用結巴斷詞(Sun, 2012)這個工具來斷詞。結巴使用概率語言模型來找出該句為最可能的詞組合。從統計資料,輸入字串分割模型的$C = C_1, C_{2,\ldots,} C_n$,和輸出$S = W_1, W_{2,\ldots,} W_m$, 其中$m \le n$。對於一個字串 C，將會有不止一種可能的分割輸出 S。分割模型的任務是找到概率最高的輸出 S 集合。

$$\mathrm{Seg(c)} \operatorname*{argmax}_{S\in G} P(S|C) = \operatorname*{argmax}_{S\in G} \frac{P(C|S)P(S)}{P(C)}$$

舉例來說:

C: 青蛙紛紛又再度回到水裡

S1: 青蛙 / 紛紛 / 又再 / 度回 / 到 / 水裡

S2: 青蛙 / 紛紛 / 又 / 再度 / 回到 / 水裡

S1 和 S2 便是不同的可能詞組合，根據不同的分割 S，詞數 m 的個數不同。m 越大，實際觀測的可能詞組成概率越小。

2.2.3 中文文字轉向量 (Chinese Word2Vec)

與語音信號和視訊訊號不同，文本是離散的，不是連續可微的。為了使它能夠輸入神經網路，成為更好的進一步分析的表示，我們使用文字向量方法將我們語料庫中的每個詞投射到低維空間的特定協調上。因此，我們現在把離散文本轉換成連續可微的詞向量。

文字向量空間(Mikolov, Sutskever, Chen, Corrado & Dean,2013)是一種資料驅動的學習表示形式，是從一個龐大的通用語料庫中學習的。在學習過程之後，模型將每個單詞

投射到隱藏的維向量空間模型中的座標上。我們可以通過計算不同詞間的協調距離來發現每個詞之間的關係。Mikolov 提出了兩種不同的方法，一是連續的詞袋(CBOW)，另一種是 Skip-gram (Skip-gram)。他還提出了兩種計算效率高的近似層次化和負採樣。利用這兩種有效的逼近方法，該模型可以以一種更有效的方式學習正確的表示方法(Goldberg & Levy, 2014)。

這裡我們使用 CBOW 來構建我們的詞嵌入網路。主要思想是利用 2c 相鄰詞的上下文向量來預測目標詞，模型結構包括 K 個詞彙量字典、輸入層、投影層和輸出層。輸入層的輸入是每個相鄰字的獨熱編碼(one-hot)向量。我們使用 K x V 輸入層維度得到每個相鄰詞上下文向量 w，w 的維度為自訂大小，而後我們得到這些上下文的平均向量 V(w)傳播到輸出層預測目標詞 w。最終，該模型可以使用反向傳播來更新參數，本篇實驗使用資料庫一以及資料庫二的全部文檔訓練出文字轉向量的網絡模型，而詞頻數過低的項目會被轉成 OOV(out of vocabulary)之向量。

2.2.4 長短期記憶神經網絡 (Long short-term memory neural network)

我們的語意流暢度需要將說故事樣本，投射到長短期記憶神經網絡後，提取遺忘閘的輸出。因此在本節中，我們將首先簡要地描述長短期記憶神經網絡(Hochreiter & Schmidhuber, 1997)。長短期記憶神經網絡是一種定向時間序列神經網路。長短期記憶神經網絡的核心是其中包含的資訊狀態更新參數$\widetilde{C_t}$:

$$\widetilde{C_t} = tanh(W_C \cdot [h_t, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t}$$

使用長短期記憶神經網絡建模的好處在於它能夠調節長期或短期記憶體上下文中保留的信息量。調節機制是使用閘的結構來完成的，閘的結構被表述為權重矩陣和啟動函數。每個長短期記憶神經網絡有三個閘(f_t，i_t，o_t)：

$$f_t = \sigma\,(W_f \cdot [h_{t-1}, x_t] + \mathrm{b}_f)$$
$$i_t = \sigma\,(W_i \cdot [h_{t-1}, x_t] + \mathrm{b}_i)$$
$$o_t = \sigma\,(W_o \cdot [h_{t-1}, x_t] + \mathrm{b}_o)$$

在細節中,忘記閘f_t,閘口負責控制過去的資訊被允許通過主要細胞$\widetilde{C_t}$的通過比率,我們利用此一機制來當作導出資訊流暢度的一種測量方法。我們首先使用監督式學習預訓練在四種故事類型的標籤上,作為初始化的模型參數,此長短期記憶網絡層的隱藏維度為 128 維。

2.2.5 語意流暢度表徵 (Lexical Coherence Representation)

為了從資訊中導出流暢度的表達特徵，我們首先使用資料集二四類童話故事，訓練出來的模型權重，作為我們預訓練好的初始化參數。然後我們用我們預訓練過的模型通過分享模型參數，去對資料集一去進行適性化的微調，這是一種轉移學習的方式。經過對資

料集一進行微調，得到最後的結果。最後提取所學習的遺忘閘矩陣W_f，提取後，我們可以計算出，在啟動函數之前，沒有偏倚項的遺忘閘的輸出值$\overline{f_t}$:

$$\overline{f_t} = W_f \cdot [h_t, x_t]$$

對於每個資料樣本，即 n 個句子,每個句子分別$\mathbf{k}_n$個字,會有總 n x k 大小的$\overline{f_t}$序列。我們將它們編碼為每個資料樣本都固定維度的表示形式。

$$F = g\,(\overline{f_1}, \ldots, \overline{f_T})$$

其中 g 表示 17 個統計泛函數，最大值(MAX)，最小值(Min)，平均數(Mean)，中位數(Medium)，標準差(Std)，1 百分位數(1%)，99 百分位數(99%)，99 百分位數- 1 百分位(99%-1%)數，偏度(skewness)，峰度(kurtosis)，最小位(min pos)，最大位(max pos)，下四分位數(lower quartile)，上四分位數(upper quartile)，上四分位數範圍(interquartile range)，冪(power)，1 差分(point difference)。這個 F 便是每個資料樣本的語意流暢度表徵。

2.3 暢度矩陣 (Cohesion Matrix)

圖 2. 流暢度矩陣特徵擷取網站示意圖
[Figure 2. Extraction procedure of cohesion matrix]

流暢度矩陣被認為是對文本和話語最詳盡的自動評估之一。它計算書寫的一致性，並能用在度量口語和敘寫的內容(McNamara *et al.*, 2014)。測量方法包括相鄰句間的相似度計算、相鄰句或整篇文章中名詞重疊(重複)、文章句子結構的相似度等。例如，相鄰句間的相似度通過每句的 LSA(潛語義分析)向量計算。然後計算兩個 LSA 向量的餘弦相似度。

圖 2 為流暢度矩陣的使用介面，將欲計算流暢度之文章文字內容輸入後，便可勾選各項特徵，得到各特徵值，例如圖中勾選的左邊便是透過 LSA 向量計算語詞間語義的相似度，而右邊則是單純計算不同詞性字詞重複的程度。

我們使用所有的 20 維特徵，通過相關性特徵選擇，我們在選擇 25%的特性集時獲得最佳性能。所選擇的五個特徵是局部名詞重疊、全功能變數名稱詞重疊、局部句子結構相似、全域句子結構相似以及句子中最小的單詞重複頻率。

局部名詞重疊是指計算相鄰句子中同一名詞的重複頻率，全功能變數名稱詞重疊是指計算同一名詞在一篇文章中千詞的重複頻率。局部和全域句子結構的相似性是指計算相鄰句子結構的相似性，計算整個句子結構的相似性。句子結構的相似性一般採用計算句子之間的編輯距離的方法(Ji & Eisenstein, 2013; Cheng & Liang, 2005)。編輯距離通常定義為將源句編輯為目標句(包括插入、刪除或切換)的成本步驟。最後一個特徵是每個句子中單詞重複頻率的平均值。

2.4 頻逆文檔頻率(TF-IDF)

方法主要由詞頻（TF）以及逆向文檔頻率（inverse document frequency，IDF）所組成，其數學式子表達成:

$$\mathrm{TF}_{i,j} = \frac{c_{i,j}}{\sum_k c_{i,j}}$$

分子是第 i 個單詞在第 j 個文章的單詞計數，分母是第 i 個單詞在全部文章中的總出現次數。

IDF 值與目標詞在整個語料庫中的通用性有關。

$$\mathrm{IDF}_i = \frac{|D|}{1 + |\{d \in D : t \in d\}|}$$

分母是整個語料庫的文章數量，分母是目標詞出現的文章數量。如果目標詞沒有出現在語料庫中，分母將為零。常用的方法是調整分母為$1 + |\{d \in D : t \in d\}|$。

最終得到值的式子如下:

$$TF-IDF_{t,d} = TF_{t,d} \times IDF_t$$

3. 實驗設計及結果 (Experimental Setup And Results)

3.1 實驗設計 (Experiment Setup)

在這篇研究中，我們使用我們提出的語意流暢度特徵表示在典型孩童與自閉症孩童之間進行分類。我們比較下列方法:

(1) 詞頻逆文檔頻率:共 2034 維的特徵值

(2) 流暢度矩陣:共 20 維的特徵值

(3) 長短期記憶神經網絡:取出隱藏層的輸出來當作特徵進行分類

(4) 語意流暢度表徵:為本篇所提出之表徵

實驗中使用了各項特徵作為流暢度特徵的比較,如 Coh-Metrix 主要是利用句子之間的相似度，來衡量一篇文章的流暢，所使用的概念是，當一篇文章裡相鄰的句子語義空間相似度高時，可以代表敘述連貫，反之則是不流暢。而透過 TF-IDF 的方法有機會讓模型透過關鍵詞數去學習,是否一篇文章一直使用過多重複的詞語,造成不流暢的現象，而 LSTM 則是在依序閱讀過段落文字之後,給出一個斷定流暢類或不流暢類的分類模型。我們提出的特徵則是保留了資料的個時間點的語義特徵分布情形，我們將這樣有各資料整個時間點語義分布的向量作為特徵，讓分類器去學習語義時序分布與流暢度的關聯。

圖 1 為以語意流暢度表徵為特徵之實驗架構圖，實驗所使用的分類器皆是支援向量機(SVM)和相關性的特徵選擇，差別僅在於使用來作為辨識基準的特徵抽換成不同的各項特徵。由於每一個兒童敘事包括不同數量的資料樣本(句數)，為了對於單個受試者給出一個單一的標籤，我們會對預測進行投票來決定最終個人的標籤。而準確率的驗證則採用交叉驗證，使用的評量方法是非加權平均召回率(UAR)。

再者，我們也利用我們提取出的流暢度特徵去對自閉症觀察診斷量表中的評分細項去做高低分分類，其中包括刻板的使用單字或片語程度、會話流暢性，以及報告事件的能力。

3.2 實驗結果 (Experiment Result)

首先，觀察特徵選擇的結果，在 TFIDF 方法方面顯示連接詞、副詞和常見的短語在漢語中，如“然後”、“還好”、“這樣”、“什麼”都是自閉症孩童的敘述重點頻率明顯高於典型孩童。關於故事內容的幾個關鍵字，例如“屋頂”、“老奶奶”、“地板”、“住”，恰恰相反，於 TD (Typically developing)受試者的敘述頻率中顯著來的高。此外，流暢度矩陣的重要特徵是“名詞重複”相鄰的句子“和”名詞是整體的重複文章“相鄰句的句子相似性”。“整篇文章的句子相似”。而這些發現，確實符合自閉症患者使用較多虛詞，較無法表達流暢的口語的情況，也印證過去的研究得到的結果。

表 1. 不同特徵訓練模型準確率
[Table 1. Performance of different features]

Features	UAR/p value	Accuracy
Coh-Metrix	0.73/0.12	0.74
TFIDF	0.77/0.10	0.78
TFIDF + Coh-Metrix **	0.80/0.04	0.79
LSTM **	0.85/0.03	0.86
Lexical Coherence Representation **	0.92/0.007	0.91

在分類準確度方面，表 1 總結我們的實驗結果。獲得的最佳準確度是通過使用我們提出的語意流暢度表徵，達到了 UAR 92%。它比其他比較的特徵流暢度矩陣和 TFIDF 方法的 73%和 77%都要出色，且是顯著來的高。而且，有趣的一點是當與使用長短期記憶模型直接執行分類的特徵擷取，去進行比較時，使用我們所提出的語意流暢度特徵所訓練的模型,在辨識典型孩童與自閉症孩童文本內容的任務上還能有 8%的召回率優勢，我們所提出的語意流暢度特徵是由長短期記憶模型的內部參數衍生而來的，而其似乎比使用整個長短期特徵模型提取的特徵更具有指標性。最後，對於不同句子數量來做資料擴充對於準確率的影響，表 2 列出了實驗的結果。資料擴充步驟中選定的使用句數，最佳數量似乎在 n=5 左右時會得到最好的效果。

表 2. 不同資料擴增參數下的準確率
[Table 2. Performance of different augment set]

# of sentence	LSTM	Lexical Coherence
n=1	0.76	0.78
n=3	0.85	0.87
n=5	0.78	0.92
n=7	0.79	0.81
n=9	0.71	0.81

而在對自閉症診斷觀察量表(ADOS)有關口語能力的評分細項中(表 3)，我們發現此篇所提出的語意流暢度特徵，的確與評分細項中跟流暢度最相關的”會話流暢性”最相關，在高低分兩類分類的任務中能達到 80.48%的準確率。

表 3. 流暢度相關行為屬性
[Table 3. Attributes of behavior related to coherence]

編號	細項內容	高低分類準確率
A4	刻板的使用單字或片語	0.5378
A8	會話流暢	0.8048
A7	報告事件	0.4208

3.3 實驗分析 (Analysis)

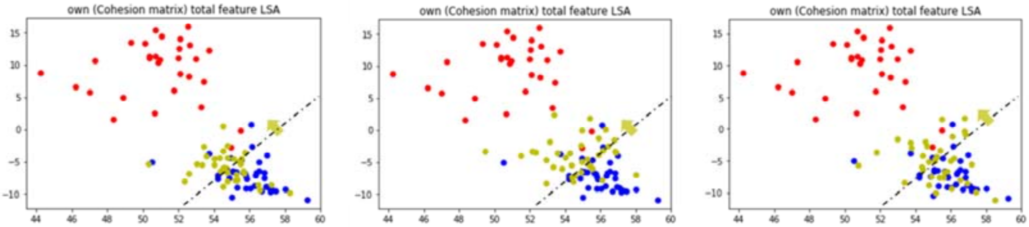


*圖 **3.** 由左至右分別為打亂字序、句子順序、刪除連貫句子中特定字詞*
[Figure 3. From left to right, the green dots represent the samples of random word order, random sentence order, delete random word in sentences]

我們的語意流暢度特徵模型實現了高識別準確性，但由於模型的複雜性，它缺乏了特徵的直接解釋性。因此，我們採用了類最近發表於 ICLR 的類似方法(Li *et al*., 2016; Koh & Liang, 2017)來進一步瞭解分析我們的特徵。在這個分析中，我們首先通過隨機打亂典型小孩資料樣本中的單詞順序、句子順序，或是刪除連貫句中的隨機詞語，模擬出語無倫次的樣本，利用潛在語義分析來降維將特徵表示成二維圖，來觀察視覺化資料分布。然後我們進一步同時使用兩種我們上述所模擬的不流暢機制(同時打亂字序或句序、同時刪除特定字與打亂句序)，來觀察不同程度的不流暢情形在我們的空間中的資料分布狀況。

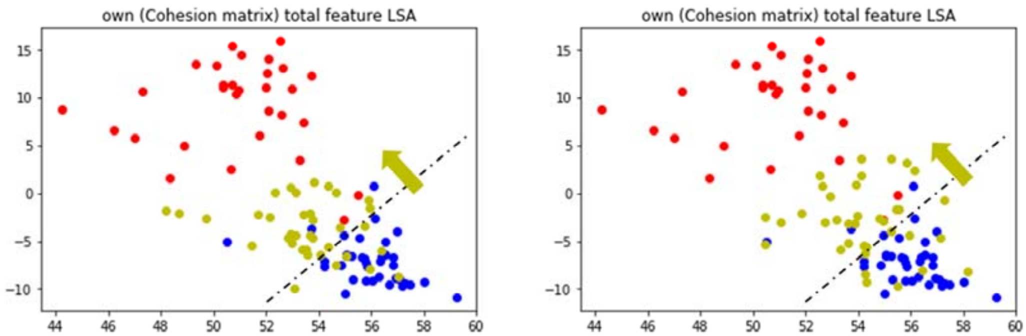


*圖 **4.** 由左至右分別為同時打亂字序或句序、同時刪除特定字與打亂句序*
[Figure 4. From left to right, the green dots represent the samples of random word and sentence order, delete random word and random sentence order]

藍點表示典型孩童的資料樣本，紅點代表自閉症孩童的資料樣本，黃點代表典型孩童故事的樣本通過模擬敘述不流暢的機制改變後的資料分布。觀察圖 4 能有趣的發現，黃色的點在典型孩童與自閉症孩童之間，顯示我們的表徵能反映出我們上述三種不流暢機制的情況。此外，隨著我們引入更多的不流暢機制(同時打亂字序或句序、同時刪除特定字與打亂句序)進入典型孩童樣本，黃點會更靠近不流暢的自閉症樣本。這個實驗分析表明，我們提出的特徵能夠反映不同的流暢機制，並且衡量文本不同的流暢程度。

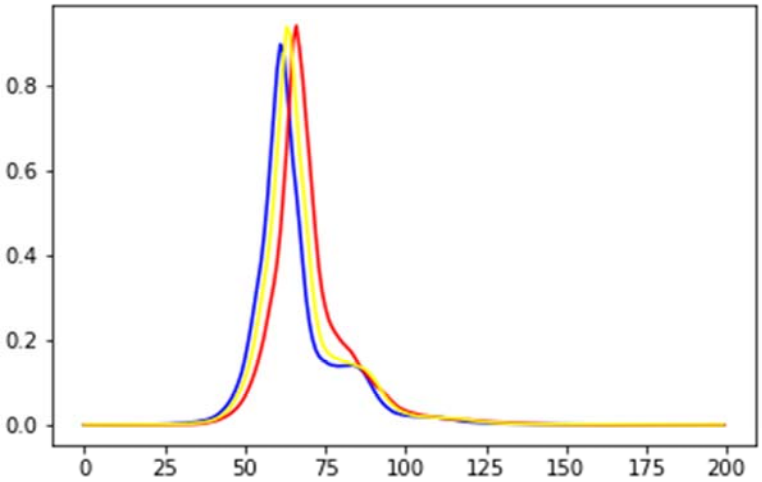


圖 5. 樣本值分布圖
[Figure 5. histogram of samples]

我們也畫出自閉症樣本、典型孩童樣本，以及導入不流暢機制的典型孩童樣本的值分布圖，從圖 5 能發現，導入不流暢機制的典型孩童樣本，的確在資料分布上更趨近於不流暢的自閉症孩童樣本。進一步我們也仔細比較各樣本中的統計值，並畫成樣本值分布圖來觀察資料分布的情況，在最大值(Max)、偏度的樣本值分布圖中，我們能明顯看出黃色的資料分布從原藍色分布區域移往紅色的資料分布區域。

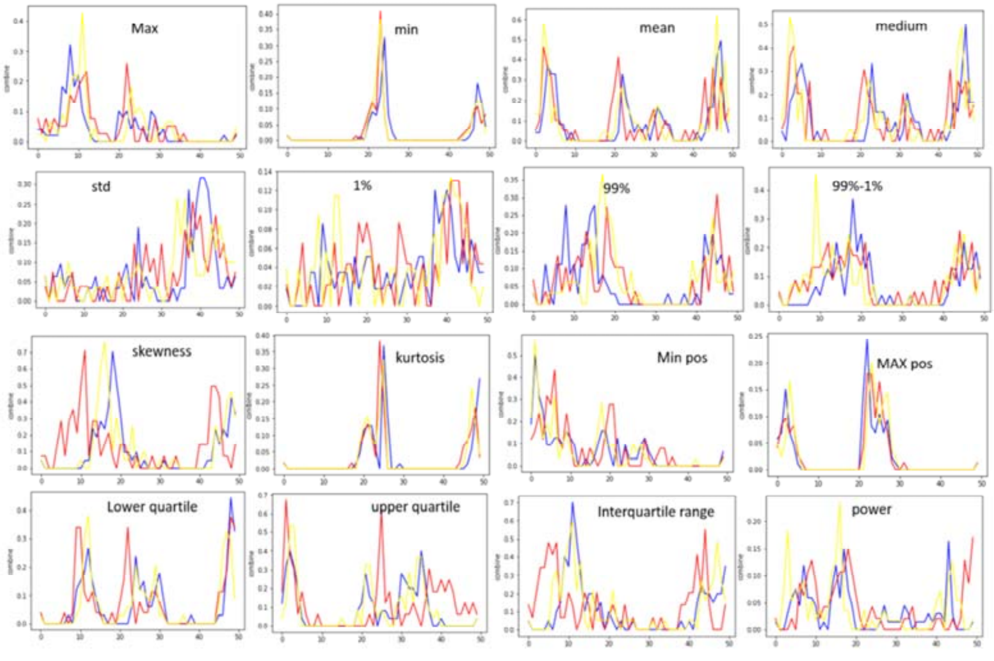


圖 6. 樣本各統計值分布圖
[Figure 6. histogram of samples with statistical analysis]

4. 結論 (Conclusions)

本文提出了一種新的資料導向的語意流暢度特徵學習架構，其精隨是利用長短期記憶神經網絡架構內的遺忘閘中，資訊隨時間傳導的機制，模擬人類閱讀時會對不同時間點資訊有不同權重的特性，來推算出流暢度特徵。用我們所提出的語意流暢度訓練的模型，在自閉症辨識能達到 92%的識別準確率，證明優於傳統衡量詞彙內容和連貫性的方法。最後,透過在典型孩童的資料樣本中導入類似不流暢的特徵(類比語無倫次的情況)，我們可以從視覺化後的資料樣本分佈點圖觀察到，將資料樣本映射到語意流暢度特徵的空間後，被導入不流暢特徵的典型小孩樣本點會移動得更加接近於自閉症小孩樣本。這樣的結果推論，即使我們的長短期記憶模型架構並非直接對流暢度標籤做學習，但在其內部架構的元素中似乎含有代表語意連貫性的流暢度的元素，甚至還能實現較高的準確率。

而在未來發展方向，首先是能去分析我們所導出的語意流暢度與計算語言學中設計的特徵之間的關係，若能找到與計算語言學中設計的特徵如語意、關鍵字、語法重複這些特徵之間的關係，便能更加瞭解神經網絡理解文章的方式。此外，流暢度也可以往多模態的方向發展，如語調流暢和手勢協調或臉部、肢體動作方面。最後未來可以套用更新的架構例如 Attention-based LSTM，並應用在現實世界的行為資訊分析，提高臨床價值將繼續是研究的中心目標。

參考文獻 (References)

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P. & Narayanan, S. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1162-1177. doi: 10.1044/2014_JSLHR-S-13-0062

Bone, D., Lee, C.-C., Chaspari, T., Gibson, J., & Narayanan, S. (2017). Signal processing and machine learning for mental health research and clinical applications [perspectives]. *IEEE Signal Processing Magazin*e, *34*(5), 195-196.

Capps, L., Losh, M., & Thurber, C. (2000). The frog ate the bug and made his mouth sad: Narrative competence in children with autism. *Journal of abnormal child psychology*, *28*(2), 193-204. doi: 10.1023/A:1005126915631

Cheng, S.-Y. & Liang, T. (2005). Chinese sentence similarity computing and appling. In *Proceedings of the 17th Conference on Computational Linguistics and Speech Processing*, 113-124.

Chorianopoulou, A., Tzinis, E., Iosif, E., Papoulidi, A., Papailiou, C., & Potamianos, A. (2017). Engagement detection for children with autism spectrum disorder. In *Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, 5055-5059. doi: 10.1109/ICASSP.2017.7953119

Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of highfunctioning children with autism spectrum disorders. *Journal of abnormal child psychology, 34*(1), 83-98. doi: 10.1007/s10802-005-9003-x

Fadaee, M., Bisazza, A., & Monz, C. (2017). Data augmentation for low-resource neural machine translation. Retrieved from arXiv preprint arXiv:1705.00440.

Goldberg, Y. & Levy, O. (2014). word2vec explained: deriving mikolov *et al*.'s negative-sampling wordembedding method. Retrieved from arXiv preprint arXiv:1402.3722.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, *36*(2), 193-202. doi: 10.3758/BF03195564

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., …Ng, A. Y. (2014). Deep speech: Scaling up end-to-end speech recognition. Retrieved from arXiv preprint arXiv:1412.5567.

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735

Ji, Y. & Eisenstein, J. (2013). Discriminative improvements to distributional sentence similarity. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 891-896.

Koh, P. W. & Liang, P. (2017). Understanding blackbox predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning (PMLR 70)*, 1885-1894.

Li, J., Monroe, W., & Jurafsky, D. (2016). Understanding neural networks through representation erasure. Retrieved from arXiv:1612.08220.

Liu, W., Li, M. & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, *9*(8), 888-898. doi: 10.1002/aur.1615

Lord, C., Rutter, M., Dilavore, P. C., & Risi, S. (2008). *ADOS: Autism diagnostic observation schedule*. Boston, MA: Hogrefe.

Losh, M. & Gordon, P. C. (2014). Quantifying narrative ability in autism spectrum disorder: A computational linguistic analysis of narrative coherence. *Journal of autism and developmental disorders*, *44*(12), 3016-3025. doi: 10.1007/s10803-014-2158-y.

Marchi, E., Schuller, B., Baron-Cohen, S., Golan, O., B¨olte, S., Arora, P., & H¨ab- Umbach, R. (2015). Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages. In *Proceedings of Sixteenth Annual Conference of the International Speech Communication Association (InterSpeech 2015)*, 115-119.

McNamara, D., Graesser, A., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511894664

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the*

26th International Conference on neural information processing systems(NIPS'13), 3111-3119.

Narayanan, S. & Georgiou, P. G. (2013). Behavioral signal processing: Deriving human behavioral informatics from speech and language. In *Proceedings of the IEEE*, *101*(5), 1203-1233.

Regneri, M., & King, D. (2016). Automated Discourse Analysis of Narrations by Adolescents with Autistic Spectrum Disorder. In *Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning*, 1-9. doi: 10.18653/v1/W16-1901

Rouhizadeh, M., Prud'Hommeaux, E., Roark, B., & Van Santen, J. (2013). Distributional semantic models for the evaluation of disordered language. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)conference*, 2013, 709-714.

Rouhizadeh, M., Sproat, R., & Van Santen, J. (2015). Similarity measures for quantifying restrictive and repetitive behavior in conversations of autistic children. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, 117-123.

Sun, J. (2012). Jiebachinese word segmentation tool. Retrieved from https://github.com/fxsjy/jieba

Tager-Flusberg, H. (1995). Once upon a Ribbit: Stories Narrated by Autistic Children. *British Journal of Developmental Psychology*, *13*(1), 45-59. doi: 10.1111/j.2044-835X.1995.tb00663.x

Computational Linguistics and Chinese Language Processing
Vol. 23, No. 1, June 2018, pp. 35-48

整合個人化磁振造影深度神經網路之演算法技術

Joint Modeling of Individual Neural Responses using a Deep Voting Fusion Network for Automatic Emotion Perception Decoding

謝宛庭*、李祈均*

Wan-Ting Hsieh and Chi-Chun Lee

摘要

不論在腦神經科學領域中，抑或是情感計算領域裡，理解聲音情感感知的潛在神經感知機制仍然是一個重要的研究方向。現今許多研究已顯示受試者中表現出的 fMRI 信號的大幅變動是由於個體差異的影響，即受試者間的變異性。然而，相對較少的研究開發用於處理此種特性的自動神經感知解碼任務中的建模技術。本研究中，我們通過學習在融合層應用的調整後的權重矩陣，提出了一種新的深度投票融合神經網絡結構的計算方法。該框架在四級聲音情緒狀態解碼任務中達到了 53.10%的 UAR 準確率，即相對於兩階段 SVM decision score 融合的改善 8.9%。此外我們嘗試音檔與 fMRI 資料的融合，藉由兩者資訊互通使得準確率可提升至 56.07%。我們的框架不僅證明了其處理個體差異的有效性，我們還加入音檔資訊，使得難過情緒分類結果大幅提升。

關鍵詞：個體差異、功能性磁振造影、聲音情緒認知、深度投票融合神經網路

Abstract

In the era underlying grouping life, affective computing and emotion recognition are closely bonding with daily life, and impose great impact on social ability. Understanding the individual differences is significant factor that should not be ignore in fMRI analysis while most of the brain studies on fMRI seldom truly deal

* 國立清華大學電機所
Department of Electrical Engineering, National Tsing Hua University
E-mail: stv7879694@gapp.nthu.edu.tw; cclee@ee.nthu.edu.tw

with it, we carry out a system considering individual variability to recognize the emotion to the vocal stimuli with BOLD signal. In our work, we propose a novel method using multimodal fusion in a voting DNN framework, where we utilize a mask on weight matrix of fusion layer to learn an individual-influenced weight matrix and realize voting in this network, and achieve 53.10% in UAR for a four-class emotion recognition task. Our analysis shows that the multimodal voting net is an effective neural network encoding individual differences and thus enhances the ability to emotion recognition. Further the join of audio feature also boosts the result to 56.07%.

Keywords: Individual Difference, fMRI, Vocal Emotion, Perception, Deep Voting Fusion Neural Net

1. 緒論 (Introduction)

現今功能性磁振造影的技術(functional Magnetic Resonance Imaging, fMRI)使得我們得以利用其擷取的血氧濃度相依對比訊號(Blood oxygen-level dependent, BOLD)觀察大腦中複雜的情緒認知機制(Zhou, Wang, Zou, Zhou & Qian, 2013; Fossati *et al*., 2003)。其中，血氧濃度相依對比訊號中具多樣性，即便是受到相同刺激的受試者也可能產生極大不同的血氧濃度相依對比訊號，而造成此現象的原因十分複雜，但大多與人體間本質上的差異有關，這也說明為何每個人都是一個與眾不同的個體。事實上神經科學相關的研究逐漸重視個體差異的議題，舉例來說，一些過去的研究顯示若僅僅將多個受試者的功能性磁振造影資料平均則會大幅地減少與腦袋結構有關的重要資訊(Van Horn, Grafton & Miller, 2008) (MacDonald, Nyberg, Sandblom, Fischer & Bäckman, 2008) (Kanai & Rees, 2011)；此外，Canli 等人也指出存在大量個體差異且不加以處理的資料也會造成較差的辨識結果以及錯誤的分析(Canli, Sivers, Whitfield, Gotlib & Gabrieli, 2002)。另外在一份延伸的研究中，Hamann 等人證實在腦區在處理情緒上更容易受到個體差異的影響(Hamann & Canli, 2004)。

在神經科學領域中，大多數的研究探討個體差異的方法為最常見的方式是列出每個人的分析結果再畫出相關性(correlation plot)來討論個體差異，例如 Dubois 等人透過在個人層面生成相關圖，驗證科學假設和收集的 fMRI 信號的可靠性(Dubois & Adolphs, 2016)；Parasuraman 等人使用類似的方法來觀察個體差異如何影響工作記憶和決策的認知過程(Parasuraman & Jiang, 2012)。這些方法揭示了考慮個體差異的重要性。然而，在開發用於從 fMRI 數據自動解碼人類情感感知的算法的背景下（例如，Wu, Chen, Liao, Kuo & Lee, 2017; Alba-Ferrara, Hausmann, Mitchell & Weis, 2011; Schirmer & Kotz, 2006)），很少建模技術將個體差異整合至演算法中。

在此研究中，我們提出深度投票融合神經網路(Deep Voting Fusion Network, DVFN)幫助我們直接整合個體差異，以便自動地進行聲音情緒解碼。其中在此架構中，我們引入融合層用以學習個體的重要程度，接著我們將觀察到的個體權重重新置入深度投票融

合神經網路中，並進行微調(finetune)。此實驗中我們招募 18 個受試者，且每個受試者受到 251 句帶有不同情緒的語句音檔刺激，而這些情緒語句乃參考 the USC IEMOCAP 資料庫(Busso *et al.*, 2008)所設計。我們提出的架構使得在對人類 4 類情緒認知的實驗上有 53.10%的 UAR(Unweighted average recall)，與前人提出的方法，兩階段式投票技術(two-stage decision-level fusion technique (Wu *et al.*, 2017))相比，本實驗結果較其進步 8.9%。此外，我們加入音檔的的 fisher vector 與 DVFN 倒數第二層的結果融合，使得準確率更提升至 56.07%。

此研究對於解構大腦與情緒之間的關係有以下幾點貢獻：

1. 此研究提出一種新的投票融合方法用以整合個體差異。
2. 此投票融合方法對於由 fMRI 資料預測受試者所受的情緒刺激能有不錯的效果。
3. 此研究亦發現加入音檔資訊，使得難過情緒分類結果大幅提升。

接下來的第二部分將針對融合(Fusion)的方法進行文獻回顧；第三部分中我們將介紹功能性磁振造影的資料收集、情緒語句資料庫的準備方法、聲音特徵擷取，以及 DVFN 架構介紹；第四部分則包含我們整體實驗架構、結果與分析；而第五部份將總結這個實驗並提出未來的研究方向。

2. 文獻回顧 (Related Work)

在工程領域中有許多方法被用來整合多模態資料(Ayache, Quénot & Gensel, 2007)，例如 early fusion 利用特徵上的融合(fusion-level)，將不同的特徵連接(concatenate)，並輸入至一個模型中訓練；另一典型的作法為 late fusion，將各模態預測出來的決策分數 (decision score) 結合，再預測一次。此外，kernel fusion 亦為一種常見的做法，利用加權平均(weighted average)的方式結合來自不同模態的特徵值再做預測。

而在深度學習神經網路架構中，最知名的多模態融合方法(Ngiam *et al.*, 2011) 為利用跨模態的自編碼神經網路(cross modality deep autoencoder)訓練出富含多模態資訊的中間層(latent space representeation).

此研究整合傳統機器學習與神經網路的概念，將多個模態以 early fusion 的方式連接作為融合層，用以學習個體的重要程度，接著我們將觀察到的個體權重重新置入深度投票融合神經網路中，進行 finetune。

3. 研究方法 (Research Methodology)

3.1 情緒性聲音刺激設計與fMRI資料收集 (Vocal Emotion Stimuli Design and Collection)

我們參考 the USC IEMOCAP 資料庫(Busso *et al.*, 2008)設計情緒性聲音資料庫做為受試者進行功能性磁振造影時的刺激材料，此材料也曾用在觀察血氧濃度相依對比訊號與韻律特徵之間的關聯(Chen, Liao, Jan, Kuo & Lee, 2016)。此次使用的資料庫含 6 種不同的刺激分別為包含情緒正向、中性、負向以及激動程度高、中、低，而每個種類包含連續 5

分鐘的聲音語句作為受試者的刺激素材。此外這些情緒性聲音語句為 the USC IEMOCAP database 中來自同一表演者所說的話，總共是 251 句話。

3.1.1 情緒分類 (Emotion Classes)

The USC IEMOCAP database 中提供每句情緒語句的情緒標籤，在整個資料庫中共分成 8 種情緒包含難過、高興、興奮、驚訝、中性、生氣、痛苦及挫折。但由於我們只擷取裡面 251 句話，使得 8 類標籤的數量分布不均，且為了驗證本研究所提出的方法，我們參考前人的做法 (Wu *et al*., 2017)，將高興、驚訝、興奮融合成一類；而生氣、痛苦、挫折融合成另一類，因此原先的 8 類情緒標前輩融合成 4 類，如表 1 所式。

表1. 本文使用的情緒標籤分類方法
[Table 1. Summary of the original and the merged labels of the 251 utterances used in this work]

原類別	Valence 等級	Arousal 等級	數量	融合後的類別	數量
難過	負向	低	33	Class 1	33
高興	正向	高	12	Class 2	79
興奮	正向	高	64		
驚訝	正向	高	3		
中性	中性	中	69	Class 3	69
生氣	負向	高	19	Class 4	70
痛苦	負向	低	1		
挫折	負向	低	50		

3.1.2 功能性磁振造影資料蒐集與前處理 (fMRI Data Collection and Preprocessing)

我們總共招募 18 位年齡介在 20~35 歲如圖 1 所示、具有大學以上學歷的受試者。此實驗設計為 block design，原實驗目的為觀察受試著對於情緒正負向 (Valence 之正向、中性及負向) 以及情緒激動程度 (Arousal 之高、中、低) 的反應，因此每個 block 會聆聽 Valence 或 Arousal 中的一種程度之連續 5 分鐘的音檔，音檔內容即上段所敘述，為帶有情緒性的語句。而每段之間設計了 5 分鐘的休息，使受試者感官狀態回復至最初。

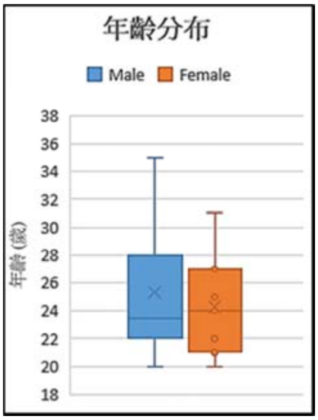


圖 1. 受試者年齡分布圖
[Figure 1. Age distribution of the subjects in this work.]

本次實驗中，我們更進一步的關注於每個 block 長達 5 分鐘的音檔裡每句話的確切情緒對於受試者的影響為何。我們使用的磁振造影掃描機台為 3T scanner (Prisma, Siemens, Germany)，TR=3(s)、體素大小(voxel size)為 3*3*3(mm^3)。接著我們利用 Data Processing Assistant for Resting-State fMRI (DPARSF)進行前處理(Yan & Zang, 2010)，此為一套基於 SPM 與 REST 開發的軟體，可由幾個簡單的按鍵自動的完成影像所有前處理的流程，包含：slice timing, realign, normalize, smooth。完成前處理後，我們將核磁共振影像內差為每秒一張以對應至情緒語句刺激的長度及語句之情緒。

3.1.3 音檔特徵擷取 (Acoustic Feature Extraction)

我們使用兩個步驟導出高維向量作為每個語句的聲音特徵：(1) 提取聲學低階描述(Low Level Description, LLD)，以及(2) 使用基於高斯混合模型（Gaussian Mixture Model, GMM）的費雪向量編碼(Fisher-vector encoding)。我們使用 Praat (Boersma, 2002)取出前 13 個 MFCC(Mel-scale 頻率倒譜係數)、音高、強度以及以 60 Hz 幀速率(framerate)提取的一階和二階特徵。我們將所有語句對中性情緒語句進行 z-score 標準化。由於每句話的長度不同，我們進一步採用 Gaussian Mixture Model- Fisher-vector encoding (GMM-FV)方法：費雪向量編碼通過首先訓練整體背景 GMM 並使用 Fisher 信息矩陣(Fisher Information Matrix, FIM)近似以進一步計算梯度向量來描述訓練的 GMM 參數所需的方向改變，即均值和方差，以獲得更好的擬合來操作關於感興趣的數據樣本，即每個話語的一系列 LLD。

3.2 利用卷積神經網路之特徵提取 (fMRI-CNN)

我們參考前人利用相同的資料庫所做的實驗與結果，發現顳葉區的腦袋資料用於聲音情緒認知辨識有最好的表現(Wu *et al.*, 2017)，此外神經科學方面研究亦顯示顳葉區確實參與許多低階聲音情緒認知作業(Phillips, Drevets, Rauch & Lane, 2003; Holt *et al.*, 2006; Schirmer & Kotz, 2006)，因此在本研究中，我們使用 Automated Anatomical Labeling (AAL)

的模板取出顳葉區的 fMRI 資料，並且利用與前人相同的手法，亦即卷積神經網路 (Convolutional Neural Network, CNN)，來抽取每個受試者顳葉區的特徵向量。

而 CNN 詳細架構如下：其中有 4 層卷積層、3 層池化層、3 層全連接層，及 1 層 softmax 輸出層以輸出對 4 類情緒的激活值，因此共為 11 層隱藏層。超參數的設定為：激活函數皆使用線性整流函數 (Rectified Linear Unit, ReLU)，優化器使用隨機梯度下降法 (Stochastic Gradient Descent, SGD)，其參數 weight decay 設定為 0.000001、momentum 設為 0.9、learning rate 為 0.0001，此外 epoch 設為 20，且訓練資料的準確率可達 88%~95% 當此深度卷積神經網路訓練完畢時，我們取出倒數第二層的隱藏層(500 個節點)作為功能性磁振造影的特徵。

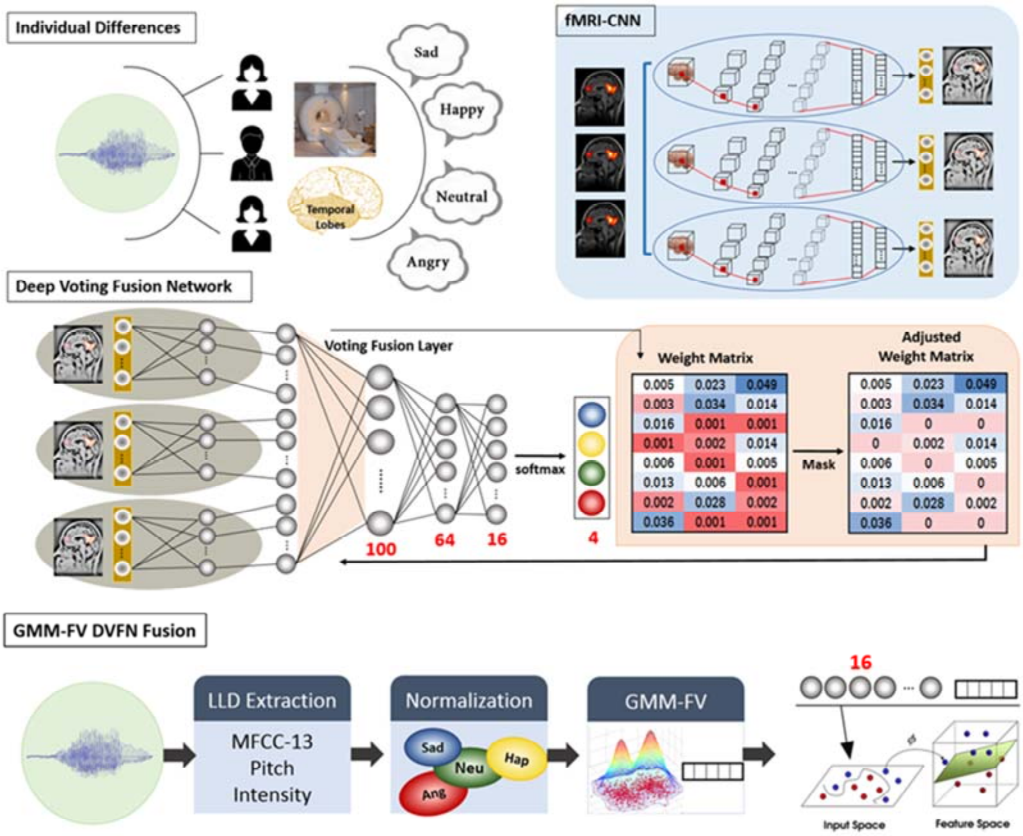


圖 2. 本文流程圖：上半部分為 fMRI 資料蒐集與特徵擷取；中間部分為 DVFN 模型架構；下半部分說明音檔特徵擷取及音檔與 DVFN 融合。
[Figure 2. The upper part of the schematic is our proposed deep voting fusion neural work in performing automatic 4-class vocal emotion decoding; While the lower part fusions the fMRI representation obtained from DVFN with the acoustic feature extracted by GMM-FV.]

3.3 深度投票融合神經網路架構 (DVFN)

我們提出的DVFN架構能幫助我們融合每個個體的特質以利認知神經影像所受的的情緒刺激。如圖2中所示，DVFN架構裡共有五層隱藏層：第一層全連階層(dense layer)將從CNN萃取出的500維特徵向量濃縮至100維；第二層為合併層，目的為將每個人的100維向量連接再一起並於第三層中投票融合降至64維再降為16維。最後為一層4個節點的softmax layer以便做情緒認知辨識。訓練DVFN時，我們使用分類交叉熵(categorical crossentropy)作為損失函數、ReLU作為激活函數、優化器選擇Adam且搭配learning rate使用0.0001，最後epoch設為10次。

3.3.1 投票融合層 (Voting Fusion Layer)

我們引入投票融合層 f 幫助我們融合多個受試者的特質，其運作方式定義如下：

$$D_f = W_f \times D_2 \tag{1}$$

其中 D_f 是融合層 f 的輸出、W_f 為融合層 f 的權重矩陣，D_2 則為 DVFN 架構中第二層的輸出，亦即每人的特徵向量合併層。值得注意的是，我們移除此層的偏差項(bias)，也不使用激活函數，使得此層聚焦於學習每個人的特徵貢獻程度(W_f)。

3.3.2 深度投票融合神經網路 (DVFN)

DVFN 架構主要進行下述兩步驟：(1)從融合層取得 W_f 後，對此權重矩陣進行遮罩(masking)，(2)將調整後的權重矩陣 $W_{adjusted}$ 重新置入融合層作為初始矩陣，並對 DVFN 進行微調(finetuning)。如上述，W_f 代表個體的貢獻程度，我們考量個體對情緒認知的能力並希望能力佳者貢獻較多、反之則抑制，因此在步驟(1)中，我們對 W_f 作遮罩，並引入一個閾值(τ)：

$$\mathrm{f(w)} = \begin{cases} 1, & if\ |w| < \tau \\ 0, & if\ |w| < \tau \end{cases} \tag{2}$$

經過遮罩後的權重強調高貢獻度的節點，且將較小的值當作擾動，如此能更有效的反應出受試者情緒認知能力的優劣。接著在步驟(2)中，我們以 $W_{adjusted}$ 取代 W_f 並重新微調(finetune)整個 DVFN 架構。

3.4 聲音特徵與fMRI特徵融合 (GMMFV-DVFN SVM-Fusion)

我們從上述的DVFN架構中取出倒數第二層(16個節點)的特徵，並將其與GMM-based 的聲音特徵 fisher vector 相接，進行 early fusion。

4. 實驗設計與結果 (Experimental Setup and Results)

我們的實驗設計為利用 fMRI 影像預測 251 個語句分別為 4 類情緒中何種情緒。在驗證訓練好的架構時，我們採用 leave-one-utterance-out 的交叉驗證方法。我們的結果皆以

unweighted average recall (UAR)顯示。以下為一些我們用來比較的架構以及其敘述，其中 AVB 代表 average-based、INB 則為 individual-based 的縮寫：

- AVB average: 直接將所有人由 fMRI-CNN 萃取出的特徵向量平均起來再透過支持向量機 (Support Vector Machine, SVM)分類。
- INB Individual: 將所有人由 fMRI-CNN 萃取出的特徵向量個別透過 SVM 分類。
- INB SVM-Voting: 將所有人由 fMRI-CNN 萃取出的特徵向量個別透過 SVM 分類並利用 decision score 投票(Wu *et al*., 2017)。
- INB DNN-Fusion: 將所有人由 fMRI-CNN 萃取出的特徵向量放入 DVFN 架構中，然而不調整節點的權重。
- INB DNN-SVM-Voting: 將所有人由 fMRI-CNN 萃取出的特徵向量放入 DVFN 架構中，並取出倒數第二層利用 SVM 分類及 decision score 投票。
- INB DVFN DNN-Fusion: 本文段落二之(三)之架構。
- INB GMMFV-SVM: 對 251 個音檔進行 fisher vector 編碼，利用 SVM 分類。
- CNN-GMMFV SVM-Fusion: 將所有人由 fMRI-CNN 萃取出的特徵向量與音檔的 fisher vector 利用 Decision score 進行 late-fusion，再利用 SVM 分類。
- DVFN-GMMFV SVM-Fusion: 將所有人由 fMRI-CNN 萃取出的特徵向量放入 DVFN 架構中取出倒數第二層的特徵，並與音檔的 fisher vector 利用 Decision score 進行 late-fusion，再利用 SVM 分類。

4.1 fMRI情緒分類結果 (Emotion Classification Results)

表 2 為我們利用 fMRI 情緒辨識的全部結果，準確率皆以 4 類 UAR 表示。我們提出的 DVFN 架構表現最佳，準確率可達 53.1%，較之前研究的結果(Wu *et al*., 2017)，亦即 INB SVM-Voting，相對進步 8.9 %。從表 2 中可以發現，將個體的特徵個別考慮，亦即 INB 與 AVB 比較，結果就有小幅提升，也因此驗證了現今腦科學研究的趨勢：個體差異需要被考量。

此外，投票神經網路，亦即投票融合層的引入，能使得神經網路學習時能共同的學習個體融合的權重，因此能更有效的提升情緒分類的結果。從表 2 中可以觀察利用神經網路投票的效果(INB DNN-Fusion)較 SVM 分類時利用 decision score 投票(INB SVM-Voting)佳；最後再加上調整權重並再次訓練 DVFN 的效果(INB DVFN DNN-Voting)則更甚於沒有調整權重時(INB DNN-Fusion)，準確率可高出 6.9%。這是因為調整權重使得神經網路能專注於貢獻度高的受試者，使他們的特徵有效的提升情緒辨識的結果。

表 2. 使用 DVFN 以及其餘融合方法的四類情緒辨識結果。準確率皆以 UAR(%) 呈現
[Table 2. It presents the 4-class emotion classification results of our proposed deep voting fusion neural network and other fusion techniques. The accuracy is measured in UAR (%).]

4-Class	***AVB: Average***	***INB: Individual***	***INB SVM-Voting***	***INB DNN-Fusion***	***INB DNN-SVM-Fusion***	***INB DVFN DNN-Voting***
Class 1	15.15	12.24	15.15	15.15	15.15	24.24
Class 2	72.15	77.43	84.81	89.87	87.34	87.34
Class 3	44.93	46.41	55.07	55.07	57.97	56.52
Class 4	37.14	40.87	40.00	38.57	47.14	44.29
UAR	42.34	44.31	48.75	49.67	51.90	**53.10**

4.1.1 閾值分析 (Threshold Analysis)

本研究中，我們還列出不同閾值下的準確率，而閾值的調整範圍為 0.001 到 0.01，並使用 Leave One person Out Cross Validation (LOOCV) 的方式進行，結果如圖 3 所示。當閾值設為 0.002 時，對於 4 類情緒預測的準確率最佳。總體來看我們發現當閾值越小(例如≦0.003)準確率看起來較高，因此若使將融合層權重中較多的值調整為 0 則效果會較好。

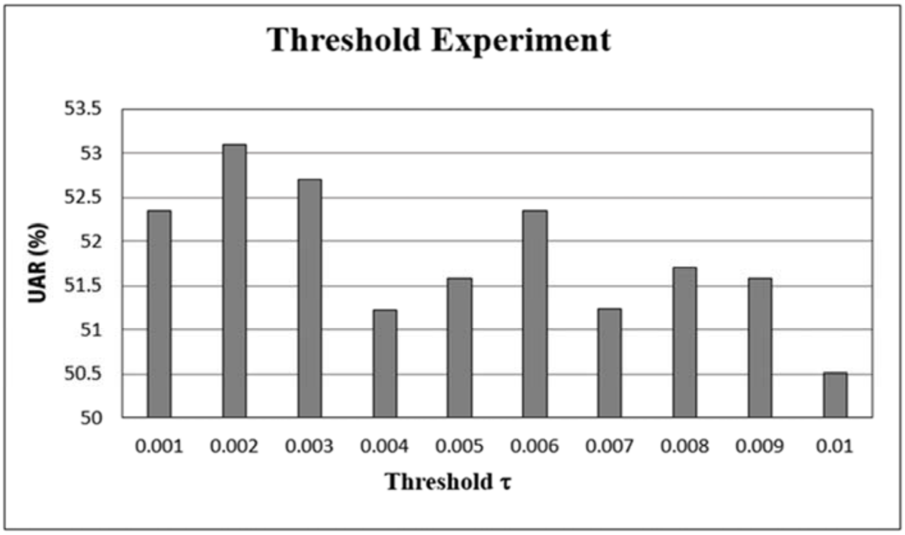


圖 3. 不同閾值 τ 下的四類情緒辨識結果。準確率皆以 UAR(%)表示
[Figure 3. The 4-class motion classification performance measured in UAR (%) as a function of threshold (τ) value]

4.2 音檔與fMRI融合結果 (Multimodal Fusion Results)

表 3 為音檔與 fMRI 融合並情緒辨識的全部結果。首先可以看到當僅利用音檔取出的GMMF特徵經過SVM對四類情緒分類(*INB GMMFV-SVM*)便可有不錯的結果(53.83 %)，且第一類的準確率高達 63.63%；相較而言利用 fMRI 資料分類時(*INB SVM-Voting*)雖準確率不差(48.75%)，但第一類的分類結果僅有 15.15%，即便在表 2 中不同實驗方法下，第一類分類最高也僅達 24.24%。由此可知音檔的特徵較易辨識難過的情緒，反觀人腦受難過情緒刺激下的變化較難辨別，因此希望藉由音檔與 fMRI 融合的方式提升第一類分類的結果，進而使整體準確率提高。

當將 CNN 抽取的特徵與音檔 GMMFV 融合時(*CNN-GMMFV SVM-Fusion*)，準確率較沒融合時佳，可達 55.25%，尤其對於第一類而言，音檔確實大幅的幫助 fMRI 資料，使其準確率達36.63%，且第四類的結果也因此種相互資訊的流通而使UAR提升至50%。此外當使用 DVFN 此種考量個體差異的模型所擷取出的 fMRI 特徵與 GMMFV 結合時(DVFN-GMMFV SVM-Fusion)，第一類的預測效果又更加進步至 42.42 %，且總準確率也較所有結果高，達到 56.07 %。

表3. 音檔與fMRI 資料融合方法的四類情緒辨識結果。準確率皆以UAR(%)呈現。[Table 3. It provides a summary of our recognition results using the fusion of audio and fMRI, and the accuracy is measured in UAR (%).]

4-Class	***INB GMMFV-SVM***	***INB SVM-Voting***	***CNN-GMMFV SVM-Fusion***	***DVFN-GMMFV SVM-Fusion***
Class 1	63.63	15.15	36.63	**42.42**
Class 2	49.37	84.81	76.08	74.68
Class 3	60.87	55.07	58.26	57.97
Class 4	41.43	40.00	50.00	49.22
UAR	53.83	48.75	55.25	**56.07**

5. 結論與未來研究 (Conclusion and Future Work)

存在於人體神經反應的個體差異在情緒認知方面以及其他高認知功能的腦區造成影響，因此對這方面的研究而言，如何處理個體差異是一項挑戰，需要以更複雜的方式模擬這項機制。我們提出創新的多人融合投票架構偵測情緒刺激，此一演算法全面的考量個人特質，透過融合層及其權重矩陣觀察個體對於預測情緒的重要程度，並藉由此深度神經網路自動化的辨識情緒刺激。利用此架構預測 4 類情緒時，我們的準確率(Unweighted average recall, UAR)可達 53.10%，此外我們發現當將權重矩陣中越多的值調整為 0 能提升準確率。且若再加入聲音資訊，準確率又可提升至 56.07%，且第一類的預測結果大幅提升 75%。

未來有許多方向可以繼續延伸，其一為利用時序架構例如遞迴神經網路(recurrent neural network, RNN)或長短期記憶神經網路(long-short term memory neural network, LSTM)模擬腦訊號並預測情緒(Li, Song, Zhang, Hou & Hu, 2017; Soleymani, Asghari-Esfeden, Fu & Pantic, 2016)，這是因為 fMRI 影像具時間資訊，可以一併放入這些專門編碼時間訊息的神經網路中以獲得更有效的特徵。其二，我們正努力延伸此概念在研究腦區與特定情緒的關聯性，期望能夠找出受情緒影響的腦區，並觀察較容易辨識情緒的受試者與不易辨識的受試者間的腦區特徵與差別。

參考文獻 (References)

ALBA-FERRARA, L., HAUSMANN, M., MITCHELL, R. L., & WEIS, S. (2011). The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. *PLoS One*, *6*(12), e28701. doi: 10.1371/journal.pone.0028701

AYACHE, S., QUÉNOT, G., & GENSEL, J. (2007). Classifier fusion for SVM-based multimedia semantic indexing. In *Proceedings of Proceedings of the 29th European conference on IR research*, 494-504. doi: 10.1007/978-3-540-71496-5_44

BOERSMA, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*(9-10), 341-345.

BUSSO, C., BULUT, M., LEE, C. C., KAZEMZADEH, A., MOWER, E., KIM, S., … NARAYANAN, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, *42*(4), 335-359. doi: 10.1007/s10579-008-9076-6

CANLI, T., SIVERS, H., WHITFIELD, S. L., GOTLIB, I. H., & GABRIELI, J. D. E. (2002). Amygdala response to happy faces as a function of extraversion. *Science*, *296*(5576), 2191. doi : 10.1126/science.1068749

CHEN, H.-Y., LIAO, Y.-H., JAN, H.-T., KUO, L.-W., & LEE, C.-C. (2016). A Gaussian mixture regression approach toward modeling the affective dynamics between acoustically-derived vocal arousal score (VC-AS) and internal brain fMRI bold signal response. In *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5775-5779. doi: 10.1109/ICASSP.2016.7472784

DUBOIS, J. & ADOLPHS, R. (2016). Building a science of individual differences from fMRI. *Trends in cognitive sciences*, *20*(6), 425-443. doi: 10.1016/j.tics.2016.03.014

FOSSATI, P., HEVENOR, S. J., GRAHAM, S. J., GRADY, C., KEIGHTLEY, M. L., CRAIK, F., …MAYBERG, H. (2003). In search of the emotional self: an fMRI study using positive and negative emotional words. *American Journal of Psychiatry*, *160*(11), 1938-1945. doi: 10.1176/appi.ajp.160.11.1938

HAMANN, S. & CANLI, T. (2004). Individual differences in emotion processing. *Current opinion in neurobiology*, *14*(2), 233-238. doi: 10.1016/j.conb.2004.03.010

HOLT, D. J., KUNKEL, L., WEISS, A. P., GOFF, D. C., WRIGHT, C. I., SHIN, L. M., …HECKERS, S. (2006). Increased medial temporal lobe activation during the passive viewing of emotional and neutral facial expressions in schizophrenia. *Schizophrenia research, 82*(2-3), 153-162. doi: 10.1016/j.schres.2005.09.021

VAN HORN, J. D., GRAFTON, S. T., & MILLER, M. B. (2008). Individual variability in brain activity: a nuisance or an opportunity? *Brain imaging and behavior*, *2*(4), 327-334. doi: 10.1007/s11682-008-9049-9

KANAI, R. & REES, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, *12*(4), 231-242. doi: 10.1038/nrn3000

LI, X., SONG, D., ZHANG, P., HOU, Y., & HU, B. (2017). Deep fusion of multi-channel neurophysiological signal for emotion recognition and monitoring. *International Journal of Data Mining and Bioinformatics*, *18*(1), 1-27. doi: 10.1504/IJDMB.2017.086097

MACDONALD, S. W., NYBERG, L., SANDBLOM, J., FISCHER, H., & BÄCKMAN, L. (2008). Increased response-time variability is associated with reduced inferior parietal activation during episodic recognition in aging. *Journal of Cognitive Neuroscience*, *20*(5), 779-786. doi: 10.1162/jocn.2008.20502

NGIAM, J., KHOSLA, A., KIM, M., NAM, J., LEE, H., & NG, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, 689-696.

PARASURAMAN, R. & JIANG, Y. (2012). Individual differences in cognition, affect, and performance: Behavioral, neuroimaging, and molecular genetic approaches. *Neuroimage*, *59*(1), 70-82. doi: 10.1016/j.neuroimage.2011.04.040

PHILLIPS, M. L., DREVETS, W. C., RAUCH, S. L., & LANE, R. (2003). Neurobiology of emotion perception I: the neural basis of normal emotion perception. *Biological psychiatry*, *54*(5), 504-514. doi: 10.1016/s0006-3223(03)00168-9

SCHIRMER, A. & KOTZ, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive sciences*, *10*(1), 24-30. doi: 10.1016/j.tics.2005.11.009

SOLEYMANI, M., ASGHARI-ESFEDEN, S., FU, Y., & PANTIC, M. (2016). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, *7*(1), 17-28. doi: 10.1109/TAFFC.2015.2436926

WU, Y., CHEN, H.-Y., LIAO, Y.-H., KUO, L.-W., & LEE, C.-C. (2017). Modeling perceivers neural-responses using lobe-dependent convolutional neural network to improve speech emotion recognition. In *Proc. of Interspeech 2017*, 3261-3265. doi: 10.21437/Interspeech.2017-562

YAN, C.-G. & ZANG, Y.-F. (2010). DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Frontiers in systems neuroscience*, *4*, 13. doi: 10.3389/fnsys.2010.00013

ZHOU, T., WANG, H., ZOU, L., ZHOU, R., & QIAN, N. (2013). A Study of Neural Mechanism in Emotion Regulation by Simultaneous Recording of EEG and fMRI Based on ICA. In: Guo C., Hou ZG., Zeng Z. (eds), *Advances in Neural Networks – ISNN 2013*.

Lecture Notes in Computer Science, (pp. 44-51). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-39065-4_6

The Association for Computational Linguistics and Chinese Language Processing

(new members are welcomed)

Aims：

1. To conduct research in computational linguistics.
2. To promote the utilization and development of computational linguistics.
3. To encourage research in and development of the field of Chinese computational linguistics both domestically and internationally.
4. To maintain contact with international groups who have similar goals and to cultivate academic exchange.

Activities：

1. Holding the Republic of China Computational Linguistics Conference (ROCLING) annually.
2. Facilitating and promoting academic research, seminars, training, discussions, comparative evaluations and other activities related to computational linguistics.
3. Collecting information and materials on recent developments in the field of computational linguistics, domestically and internationally.
4. Publishing pertinent journals, proceedings and newsletters.
5. Setting of the Chinese-language technical terminology and symbols related to computational linguistics.
6. Maintaining contact with international computational linguistics academic organizations.
7. Dealing with various other matters related to the development of computational linguistics.

To Register：

Please send application to:

The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

payment： Credit cards(please fill in the order form), cheque, or money orders.

Annual Fees：

regular/overseas member： NT$ 1,000 (US$50.-)
group membership： NT$20,000 (US$1,000.-)
life member：ten times the annual fee for regular/ group/ overseas members

Contact：

Address： The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

Tel.：886-2-2788-3799 ext. 1502 Fax：886-2-2788-1638

E-mail: aclclp@hp.iis.sinica.edu.tw Web Site: http://www.aclclp.org.tw

Please address all correspondence to Miss Qi Huang, or Miss Abby Ho

The Association for Computational Linguistics and Chinese Language Processing

Membership Application Form

Member ID#： ________________

Name： __________________________________ Date of Birth： _______________

Country of Residence： ___________________ Province/State： ___________________

Passport No.： ___________________________ Sex: ________________________

Education(highest degree obtained)： ____

Work Experience： ____

____

Present Occupation： ____

Address： ____

____

Email Add： ____

Tel. No： ____________________________ Fax No： ___________________________

Membership Category：☐ Regular Member ☐ Life Member

Date： _____/_____/_____ （Y-M-D）

Applicant's Signature：

Remarks： Please indicated clearly in which membership category you wish to register, according to the following scale of annual membership dues：

Regular Member ： US$ 50.- （NT$ 1,000）

Life Member ： US$500.-（NT$10,000）

Please feel free to make copies of this application for others to use.

Committee Assessment：

中華民國計算語言學學會

宗旨：

（一） 從事計算語言學之研究
（二） 推行計算語言學之應用與發展
（三） 促進國內外中文計算語言學之研究與發展
（四） 聯繫國際有關組織並推動學術交流

活動項目：

（一）定期舉辦中華民國計算語言學學術會議（Rocling）
（二）舉行有關計算語言學之學術研究講習、訓練、討論、觀摩等活動項目
（三）收集國內外有關計算語言學知識之圖書及最新發展之資料
（四）發行有關之學術刊物，論文集及通訊
（五）研定有關計算語言學專用名稱術語及符號
（六）與國際計算語言學學術機構聯繫交流
（七）其他有關計算語言發展事項

報名方式：

1. 入會申請書：請至本會網頁下載入會申請表，填妥後郵寄或E-mail至本會
2. 繳交會費：劃撥：帳號：19166251，戶名：中華民國計算語言學學會
   信用卡：請至本會網頁下載信用卡付款單

年費：

終身會員： 10,000.- （US$ 500.-）
個人會員： 1,000.- （US$ 50.-）
學生會員： 500.- （限國內學生）
團體會員： 20,000.- （US$ 1,000.-）

連絡處：

地址：台北市115南港區研究院路二段128號 中研院資訊所(轉)
電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
E-mail：aclclp@hp.iis.sinica.edu.tw 網址: http://www.aclclp.org.tw
連絡人：黃琪 小姐、何婉如 小姐

中華民國計算語言學學會
個人會員入會申請書

<table>
<tr><td>會員類別</td><td>☐終身 ☐個人 ☐學生</td><td colspan="2">會員編號</td><td colspan="2">（由本會填寫）</td></tr>
<tr><td rowspan="2">姓　　名</td><td rowspan="2"></td><td rowspan="2">性別</td><td rowspan="2"></td><td>出生日期</td><td>年　月　日</td></tr>
<tr><td>身分證號碼</td><td></td></tr>
<tr><td>現　　職</td><td></td><td colspan="2">學　　歷</td><td colspan="2"></td></tr>
<tr><td>通訊地址</td><td colspan="5">□□□</td></tr>
<tr><td>戶籍地址</td><td colspan="5">□□□</td></tr>
<tr><td>電　　話</td><td></td><td colspan="2">E-Mail</td><td colspan="2"></td></tr>
<tr><td colspan="6">申請人:　　　　　　（簽章）

中　華　民　國　　　年　　月　　日</td></tr>
</table>

審查結果:

1. 年費：

   終身會員：　10,000.-
   個人會員：　1,000.-
   學生會員：　500.-（限國內學生）
   團體會員：　20,000.-

2. 連絡處：

   地址：台北市南港區研究院路二段128號 中研院資訊所(轉)
   電話：(02) 2788-3799　ext.1502 傳真：(02) 2788-1638
   E-mail：aclclp@hp.iis.sinica.edu.tw　網址: http://www.aclclp.org.tw
   連絡人：黃琪 小姐、何婉如 小姐

3. 本表可自行影印

The Association for Computational Linguistics and Chinese Language Processing (ACLCLP)
PAYMENT FORM

Name: ________________________________(Please print) Date: ______________

Please debit my credit card as follows: US$ ____________

❑ VISA CARD ❑ MASTER CARD ❑ JCB CARD Issue Bank:_________________

Card No.: ________ -_________-________ -_________ Exp. Date:___________(M/Y)

3-digit code: _________ (on the back card, inside the signature area, the last three digits)

CARD HOLDER SIGNATURE: ________________________________

Phone No.: ________________________E-mail: _________________________________

Address: ____

PAYMENT FOR

US$ _____ ❑ Computational Linguistics & Chinese Languages Processing (IJCLCLP)

Quantity Wanted: ______

US$ _____ ❑ Journal of Information Science and Engineering (JISE)

Quantity Wanted: ______

US$ _____ ❑ Publications:______________________________

US$ _____ ❑ Text Corpora: ____________________________

US$ _____ ❑ Speech Corpora:___________________________

US$ _____ ❑ Others: _________________________________

US$ _________ ❑ Membership Fees ❑ Life Membership ❑ New Membership ❑Renew

US$ _________ = Total

Fax 886-2-2788-1638 or Mail this form to:

ACLCLP

% IIS, Academia Sinica

Rm502, No.128, Sec.2, Academia Rd., Nankang, Taipei 115, Taiwan

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: http://www.aclclp.org.tw

中 華 民 國 計 算 語 言 學 學 會
信用卡付款單

姓名: ______________________________(請以正楷書寫)　　日期:：______________________

卡別：❑ VISA CARD　　❑ MASTER CARD ❑ JCB CARD　　發卡銀行：_______________

信用卡號：__________-__________-__________-__________　　有效日期：__________(m/y)

卡片後三碼：_______________（卡片背面簽名欄上數字後三碼）

持卡人簽名：____(簽名方式請與信用卡背面相同)

通訊地址：____

聯絡電話：______________________________E-mail：___________________________________

備註：為順利取得信用卡授權，請提供與發卡銀行相同之聯絡資料。

付款內容及金額：

NT$__________　❑ 中文計算語言學期刊(IJCLCLP) ________________

NT$__________　❑ Journal of Information Science and Engineering (JISE)

NT$__________　❑ 中研院詞庫小組技術報告______________________

NT$__________　❑ 文字語料庫 ______________________________________

NT$__________　❑ 語音資料庫 ______________________________________

NT$__________　❑ 光華雜誌語料庫1976~2010

NT$__________　❑ 中文資訊檢索標竿測試集/文件集

NT$__________　❑ 會員年費：❑續會　　　❑新會員　　　❑終身會員

NT$__________　❑ 其他: ______________________________________

NT$_______________ = 合計

填妥後請傳真至 02-27881638 或郵寄至:

11529台北市南港區研究院路2段128號中研院資訊所(轉)中華民國計算語言學學會 收

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: http://www.aclclp.org.tw

Publications of the Association for Computational Linguistics and Chinese Language Processing

		Surface	AIR (US&EURP)	AIR (ASIA)	VOLUME	AMOUNT
1.	no.92-01, no. 92-04(合訂本) ICG 中的論旨角色與 A Conceptual Structure for Parsing Mandarin -- Its Frame and General Applications--	US$ 9	US$ 19	US$15	_____	_____
2.	no.92-02 V-N 複合名詞討論篇 & 92-03 V-R 複合動詞討論篇	12	21	17	_____	_____
3.	no.93-01 新聞語料庫字頻統計表	8	13	11	_____	_____
4.	no.93-02 新聞語料庫詞頻統計表	18	30	24	_____	_____
5.	no.93-03 新聞常用動詞詞頻與分類	10	15	13	_____	_____
6.	no.93-05 中文詞類分析	10	15	13	_____	_____
7.	no.93-06 現代漢語中的法相詞	5	10	8	_____	_____
8.	no.94-01 中文書面語頻率詞典（新聞語料詞頻統計）	18	30	24	_____	_____
9.	no.94-02 古漢語字頻表	11	16	14	_____	_____
10.	no.95-01 注音檢索現代漢語字頻表	8	13	10	_____	_____
11.	no.95-02/98-04 中央研究院平衡語料庫的內容與說明	3	8	6	_____	_____
12.	no.95-03 訊息為本的格位語法與其剖析方法	3	8	6	_____	_____
13.	no.96-01 「搜」文解字－中文詞界研究與資訊用分詞標準	8	13	11	_____	_____
14.	no.97-01 古漢語詞頻表（甲）	19	31	25	_____	_____
15.	no.97-02 論語詞頻表	9	14	12	_____	_____
16.	no.98-01 詞頻詞典	18	30	26	_____	_____
17.	no.98-02 Accumulated Word Frequency in CKIP Corpus	15	25	21	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	4	9	7	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	8	13	11	_____	_____
20.	Computational Linguistics & Chinese Languages Processing (One year) (Back issues of *IJCLCLP*: US$ 20 per copy)	---	100	100	_____	_____
21.	Readings in Chinese Language Processing	25	25	21	_____	_____
				TOTAL	_____	_____

10% member discount: ___________Total Due:__________

- **OVERSEAS USE ONLY**
- PAYMENT： ☐ Credit Card (Preferred)
  ☐ Money Order or Check payable to "The Association for Computation Linguistics and Chinese Language Processing " or “中華民國計算語言學學會”
- E-mail：aclclp@hp.iis.sinica.edu.tw

Name (please print): ____________________________ Signature: ____________________________

Fax: ____ E-mail: ______________________________

Address：____

中華民國計算語言學學會
相關出版品價格表及訂購單

編號	書目	會　員	非會員	冊數	金額
1.	no.92-01, no. 92-04 (合訂本)　ICG 中的論旨角色 與 A conceptual Structure for Parsing Mandarin--its Frame and General Applications--	NT$ 80	NT$ 100	_____	_____
2.	no.92-02, no. 92-03 (合訂本) V-N 複合名詞討論篇 與V-R 複合動詞討論篇	120	150	_____	_____
3.	no.93-01　新聞語料庫字頻統計表	120	130	_____	_____
4.	no.93-02　新聞語料庫詞頻統計表	360	400	_____	_____
5.	no.93-03　新聞常用動詞詞頻與分類	180	200	_____	_____
6.	no.93-05　中文詞類分析	185	205	_____	_____
7.	no.93-06　現代漢語中的法相詞	40	50	_____	_____
8.	no.94-01　中文書面語頻率詞典（新聞語料詞頻統計）	380	450	_____	_____
9.	no.94-02　古漢語字頻表	180	200	_____	_____
10.	no.95-01　注音檢索現代漢語字頻表	75	85	_____	_____
11.	no.95-02/98-04　中央研究院平衡語料庫的內容與說明	75	85	_____	_____
12.	no.95-03　訊息爲本的格位語法與其剖析方法	75	80	_____	_____
13.	no.96-01「搜」文解字－中文詞界研究與資訊用分詞標準	110	120	_____	_____
14.	no.97-01　古漢語詞頻表（甲）	400	450	_____	_____
15.	no.97-02　論語詞頻表	90	100	_____	_____
16	no.98-01　詞頻詞典	395	440	_____	_____
17.	no.98-02　Accumulated Word Frequency in CKIP Corpus	340	380	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	90	100	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	75	85	_____	_____
20	論文集 COLING 2002 紙本	100	200	_____	_____
21.	論文集 COLING 2002 光碟片	300	400	_____	_____
22.	論文集 COLING 2002 Workshop 光碟片	300	400	_____	_____
23.	論文集 ISCSLP 2002 光碟片	300	400	_____	_____
24.	交談系統暨語境分析研討會講義（中華民國計算語言學學會1997第四季學術活動）	130	150	_____	_____
25.	中文計算語言學期刊（一年四期）年份：______（過期期刊每本售價500元）	---	2,500	_____	_____
26.	Readings of Chinese Language Processing	675	675	_____	_____
27.	剖析策略與機器翻譯 1990	150	165	_____	_____
			合　計	_____	_____

※　此價格表僅限國內 (台灣地區) 使用

劃撥帳戶：中華民國計算語言學學會　　劃撥帳號：19166251

聯絡電話：(02) 2788-3799 轉1502

聯絡人： 黃琪 小姐、何婉如 小姐　　E-mail:aclclp@hp.iis.sinica.edu.tw

訂購者：______________________________ 收據抬頭：______________________________

地　　址：____

電　　話：______________________________ E-mail:______________________________

Information for Authors

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) invites submission of original research papers in the area of computational linguistics and speech/text processing of natural language. All papers must be written in English or Chinese. Manuscripts submitted must be previously unpublished and cannot be under consideration elsewhere. Submissions should report significant new research results in computational linguistics, speech and language processing or new system implementation involving significant theoretical and/or technological innovation. The submitted papers are divided into the categories of regular papers, short paper, and survey papers. Regular papers are expected to explore a research topic in full details. Short papers can focus on a smaller research issue. And survey papers should cover emerging research trends and have a tutorial or review nature of sufficiently large interest to the Journal audience. There is no strict length limitation on the regular and survey papers. But it is suggested that the manuscript should not exceed 40 double-spaced A4 pages. In contrast, short papers are restricted to no more than 20 double-spaced A4 pages. All contributions will be anonymously reviewed by at least two reviewers.

Copyright：It is the author's responsibility to obtain written permission from both author and publisher to reproduce material which has appeared in another publication. Copies of this permission must also be enclosed with the manuscript. It is the policy of the CLCLP society to own the copyright to all its publications in order to facilitate the appropriate reuse and sharing of their academic content. A signed copy of the IJCLCLP copyright form, which transfers copyright from the authors (or their employers, if they hold the copyright) to the CLCLP society, will be required before the manuscript can be accepted for publication. The papers published by IJCLCLP will be also accessed online via the IJCLCLP official website and the contracted electronic database services.

Style for Manuscripts: The paper should conform to the following instructions.
1. Typescript: Manuscript should be typed double-spaced on standard A4 (or letter-size) white paper using size of 11 points or larger.
2. Title and Author: The first page of the manuscript should consist of the title, the authors' names and institutional affiliations, the abstract, and the corresponding author's address, telephone and fax numbers, and e-mail address. The title of the paper should use normal capitalization. Capitalize only the first words and such other words as the orthography of the language requires beginning with a capital letter. The author's name should appear below the title.
3. Abstracts and keywords: An informative abstract of not more than 250 words, together with 4 to 6 keywords is required. The abstract should not only indicate the scope of the paper but should also summarize the author's conclusions.
4. Headings: Headings for sections should be numbered in Arabic numerals (i.e. 1.,2....) and start form the left-hand margin. Headings for subsections should also be numbered in Arabic numerals (i.e. 1.1. 1.2...).
5. Footnotes: The footnote reference number should be kept to a minimum and indicated in the text with superscript numbers. Footnotes may appear at the end of manuscript
6. Equations and Mathematical Formulas: All equations and mathematical formulas should be typewritten or written clearly in ink. Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses.
7. References: All the citations and references should follow the APA format. The basic form for a reference looks like

Authora, A. A., Authorb, B. B., & Authorc, C. C. (Year). Title of article. *Title of Periodical, volume number*(issue number), pages.

Here shows an example.

Scruton, R. (1996). The eclipse of listening. *The New Criterion, 15*(30), 5-13.

The basic form for a citation looks like (Authora, Authorb, and Authorc, Year). Here shows an example. (Scruton, 1996).
Please visit the following websites for details.
(1) APA Formatting and Style Guide (http://owl.english.purdue.edu/owl/resource/560/01/)
(2) APA Stytle (http://www.apastyle.org/)

No page charges are levied on authors or their institutions.

Final Manuscripts Submission: If a manuscript is accepted for publication, the author will be asked to supply final manuscript in MS Word or PDF files to clp@hp.iis.sinica.edu.tw

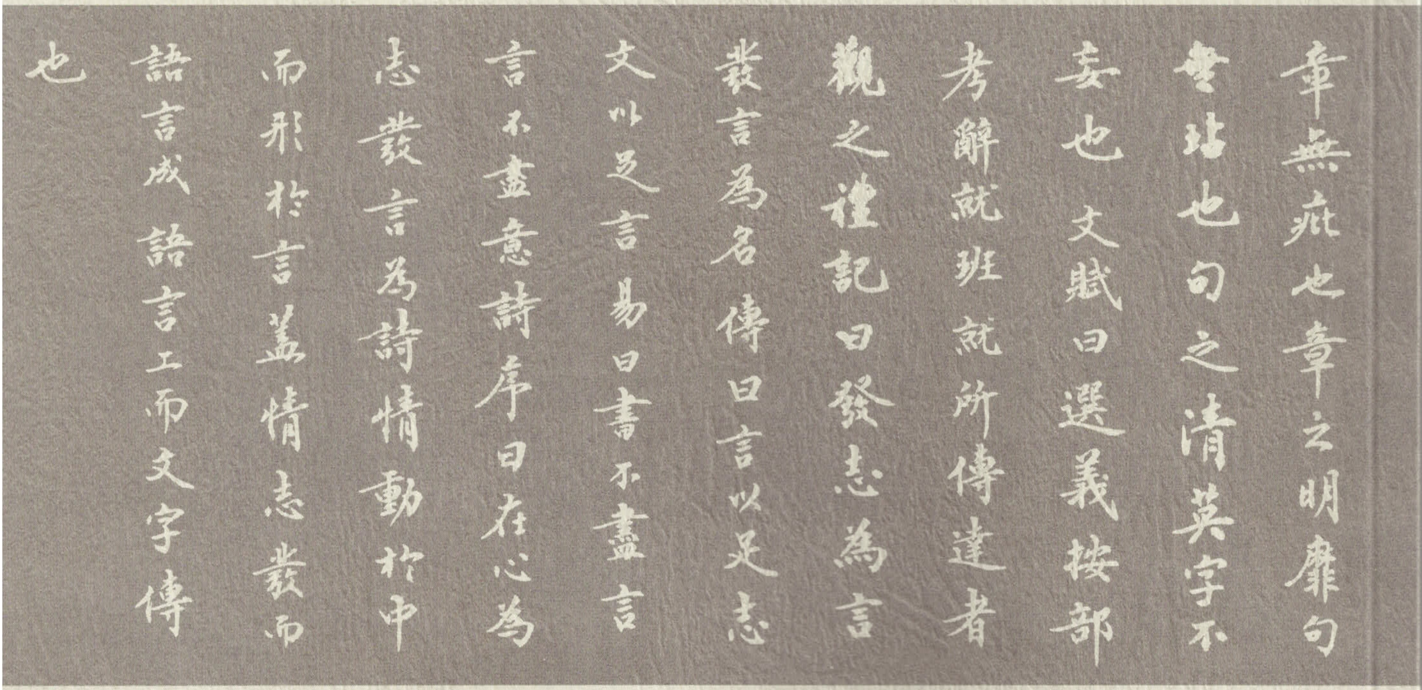
Online Submission: http://www.aclclp.org.tw/journal/submit.php

Please visit the IJCLCLP Web page at http://www.aclclp.org.tw/journal/index.php

International Journal of
Computational Linguistics & Chinese Language Processing
vol.23 no.1 June 2018

Contents

Papers



ISSN: 1027-376X
The Association for Computational Linguistics and Chinese Language Processing